NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

COMPARISON OF PROFICIENCY OBJECTIVES, PERFORMANCE OBJECTIVES, AND SUCCESS AT FOLLOW-ON TRAINING

by

Johnna M. Marchant

September 1998

Thesis Advisor:	Samuel E. Buttrey
Second Reader:	John Thain

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE September 1998	3. REPORT TYPE AND DATES COVERED Master's Thesis

4. TITLE AND SUBTITLE COMPARISON OF PROFICIENCY OBJECTIVES, PERFORMANCE OBJECTIVES, and SUCCESS AT FOLLOW-ON TRAINING	5. FUNDING NUMBERS
6. AUTHOR(S) Johnna M. Marchant	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	12b. DISTRIBUTION CODE

13. ABSTRACT *(maximum 200 words)*

The Defense Language Institute Foreign Language Center (DLIFLC) trains students in over 21 foreign languages for the Department of Defense (DoD). The National Security Agency (NSA) and Defense Intelligence Agency (DIA) are responsible for setting the training objectives for students entering professional fields in intelligence.

In the past, general proficiency in listening, reading, and speaking skills has been the focus of language learning and testing in the DoD. Certain minimum scores on the Defense Language Proficiency Test (DLPT) are required for certain training and operational positions within the DoD.

DoD has not established applicable performance objective scores for training and operational positions. Individual service commanders at DLIFLC may exercise some discretion in borderline cases where general minimum DLPT requirements have not been met. They may take into account performance objective scores and grant waivers for attending Goodfellow Air Force Base (GAFB) follow-on training.

The purpose of this study is to determine how the performance objective scores relate to success on the DLPT and how the combination of DLPT and performance objective tests might possibly relate to success on follow-on training at GAFB. Success at GAFB is defined by on-time graduation, number of required special-assistance hours, and performance on "block" tests.

14. SUBJECT TERMS DLPT, performance objective	15. NUMBER OF PAGES 89 104
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

298-102

Approved for public release; distribution is unlimited

COMPARISON OF PROFICIENCY OBJECTIVES, PERFORMANCE OBJECTIVES AND SUCCESS AT FOLLOW-ON TRAINING

Johnna M. Marchant
Lieutenant, United States Navy
B.S., Portland State University, 1986

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1998

Author: Johnna M. Marchant
Johnna M. Marchant

Approved by: Samuel E. Buttrey
Samuel E. Buttrey, Thesis Advisor

J. W. Thain
John Thain, Second Reader

Richard E. Rosenthal
Richard E. Rosenthal, Chairman
Department of Operations Research

ABSTRACT

The Defense Language Institute Foreign Language Center (DLIFLC) trains students in over 21 foreign languages for the Department of Defense (DoD). The National Security Agency (NSA) and Defense Intelligence Agency (DIA) are responsible for setting the training objectives for students entering professional fields in intelligence.

In the past, general proficiency in listening, reading, and speaking skills has been the focus of language learning and testing in the DoD. Certain minimum scores on the Defense Language Proficiency Test (DLPT) are required for certain training and operational positions within the DoD. DoD has not established applicable performance objective scores for training and operational positions. Individual service commanders at DLIFLC may exercise some discretion in borderline cases where general minimum DLPT requirements have not been met. They may take into account performance objective scores and grant waivers for attending Goodfellow Air Force Base (GAFB) follow-on training.

The purpose of this study is to determine how the performance objective scores relate to success on the DLPT and how the combination of DLPT and performance objective tests might possibly relate to success on follow-on training at GAFB. Success at GAFB is defined by on-time graduation, number of required special-assistance hours, and performance on "block" tests.

TABLE OF CONTENTS

TABLE OF SYMBOLS, ACRONYMS, AND ABBREVIATIONS

β_j, $\boldsymbol{\beta}$	A regression coefficient, or (in bold) a vector of coefficients
b_j, $\mathbf{b}$	An estimated regression coefficient, or (in bold) a vector of estimated coefficients
df	Degrees of Freedom
DLIFLC	Defense Language Institute Foreign Language Center
DLPT	Defense Language Proficiency Test
DoD	Department of Defense
ε_i, $\boldsymbol{\varepsilon}$	Regression "error" term for individual i or (in bold) a vector of errors
ESR	Research and Analysis Division, DLIFLC
FLO	Final Learning Objective
FY	Fiscal Year
GAFB	Goodfellow Air Force Base
H_0	Null hypothesis
IET	Initial Entry Trainees
k	Number of independent variables in regression
LASCP	Language Skill Change Project
n	Number of observations
NSA	National Security Agency
OLS	Ordinary Least Squares
Φ	Normal cumulative distribution function
RSS	Residual Sum of Squares
SD	Standard Deviation
SE	Standard Error
SIA	Special Individual Assistance
SPSS	Statistical Package for Social Sciences
X_i	Vector of values of independent variables for individual *I*
$\mathbf{X}$	Matrix ($n \times k$) of values of independent variables
Y_i	Value of dependent variable for individual *I*
$\mathbf{Y}$	Vector (length n) of values of dependent variable

EXECUTIVE SUMMARY

The Defense Language Institute Foreign Language Center (DLIFLC) trains students in over 21 foreign languages for the Department of Defense (DoD). The National Security Agency (NSA) and Defense Intelligence Agency (DIA) are responsible for setting the training objectives for students entering professional fields in intelligence.

In the past, general proficiency in listening, reading, and speaking skills has been the focus of language learning and testing in the DoD. Certain minimum scores on the Defense Language Proficiency Test (DLPT) are required for certain training and operational positions within the DoD.

DoD has not established applicable performance objective test scores for training and operational positions. Individual service commanders at DLIFLC may exercise some discretion in borderline cases where general minimum DLPT requirements have not been met. They may take into account performance objective scores and grant waivers for attending Goodfellow Air Force Base (GAFB) follow-on training.

The aims of the study were to determine how the performance objective scores relate to success on the DLPT and how a combination of DLPT and performance objective tests might possibly relate to success on follow-on training at GAFB. In part, we seek "cut-off" scores on performance objective tests that will correlate to success on DLPTs and at GAFB. Success at GAFB is defined by on-time graduation, number of required special-assistance hours, and performance on "block tests."

In the first phase of the study, we used stepwise multiple linear regression to create a model, which showed which performance objectives correlated best to the DLPT

score for each language. Once the models were produced, we looked for consistency in the correlation of performance objectives and the DLPT amongst all the languages, then by the category of language difficulty, and finally by category of alphabet type (either Roman or non-Roman).

We then determined cut-off scores for the performance objectives for each language that had one performance objective correlating to the DLPT. We calculated the cut-off score assuming a Normal probability distribution for DLPT scores, with mean determined by the performance objective score. The cutoff was the performance objective score that gave an 80 percent chance of passing the DLPT.

For the models that had two performance objectives correlating to the DLPT, we created a graph that given one performance objective score determines what the student needs to achieve on the second performance objective to have an 80 percent chance of passing the DLPT. A passing grade on the DLPT was a score of 40 for DLPT_L (listening) and DLPT_R (reading), and 20 for DLPT_S (speaking).

Additionally, we conducted an evaluation of the quality of the models. We looked at how well the models described the variation of the DLPT and whether or not there was a negative correlation between the performance objectives and the DLPT. The negative correlation of a performance objective and the DLPT does not make "good" sense by itself, because it states that students scoring score higher on a performance objective are expected to score lower on the DLPT. The belief is that there is a more complicated explanation that could be explained by interactions between performance objectives, but since we did not allow interactions in these models, some models show a negative correlation.

In the second phase of this study, stepwise multiple linear regression was used to determine the correlation of performance objectives and DLPT scores with "block" tests at GAFB for each service. In this phase, attention was restricted to the Russian language. We looked for consistency in the performance objectives and the DLPT to determine if there was one objective that best determined success at GAFB.

In the first phase, the study found that in some languages the performance objectives were better predictors of success on the DLPT than other languages. Polish and Japanese were languages where the performance objectives were "good" predictors for performance on the DLPT. Vietnamese was a language where the performance objectives were "poor" predictors for performance on the DLPT.

There are ten performance objective tests. Numbers 1 through 4 are intended to measure listening skills; numbers 5 through 8 are aimed at measuring reading skills; and numbers 9 and 10 measure speaking. We found that, across all languages, performance objectives 1, 3, and 7 appeared most frequently as predictors of success on the DLPT_L. Performance objectives 2, 5, and 7 were the best predictors for success on the DLPT_R. And finally, performance objective 1 was the most frequent predictor for success on the DLPT_S.

These results are slightly different when the languages are divided by categories of difficulty (I to IV, I being easiest) and by alphabet (Roman and non-Roman), but the general conclusion remains valid: the performance objective tests do not seem to measure what they were designed to measure. Furthermore, different performance objective tests appear as the "best" predictors of DLPT tests scores in different languages. For example, proficiency objective 9 was the best predictor for DLPT_L in Czech, while proficiency objective 7 was the best predictor for DLPT_L in Hebrew.

For the GAFB, again some of the proficiency tests were better predictors of success than others. The best predictors of success on the "block" tests are different for the three courses (Army, Navy/Marine Corps and Air Force).

The study shows that the performance objectives are not measuring the listening, reading, and speaking skills intended, nor do they seem to measure the same things in different languages. We recommend that DLIFLC review and validate their performance objectives. If cut-off scores for performance objectives need to be assigned, DLIFLC can assign them utilizing the findings within this thesis.

I. INTRODUCTION

A. BACKGROUND

The Defense Language Institute Foreign Language Center (DLIFLC) trains students in 21 foreign languages for the Department of Defense (DoD). The National Security Agency (NSA) and Defense Intelligence Agency (DIA) are responsible for setting the training objectives for students entering professional fields in intelligence.

In the early 1990s these two communities developed specific training objectives for students entering the basic language program. These objectives were written and refined over a period of several years with the assistance of numerous experienced personnel in the various fields that the students were preparing to enter. With NSA and DIA concurrence, DLIFLC combined the requirements from both communities into a single set of program objectives for all students. These objectives are referred to as Final Learning Objectives (FLO).

There are four types of FLOs: **proficiency objectives**, which include the general language skills of reading, listening and speaking; **performance objectives**, which focus on job-specific skills that involve foreign language use such as transcribing, summarizing text, translating, etc.; **content objectives**, which include background knowledge of the target country related to interpretation of foreign language materials --such as knowledge in the area of politics, military topics, culture, geography and technology; and **enabling objectives**, which incorporate knowledge of colloquial language, dictionary usage, number drills, and future transliteration system. Test instruments and test data are available for measuring only the first two kinds of FLOs, proficiency objectives and performance objectives. This study will be concerned only with data on these two FLOs.

DLIFLC measures attainment of proficiency FLOs through the Defense Language Proficiency Tests (DLPT) and the performance objectives through ten performance objective tests. Since 1958 various formats and scoring systems have been used in different versions of the DLPT to measure general language proficiency. The current DLPT consists of two multiple-choice tests and an interview. The multiple-choice tests measure proficiency in listening and reading and the interview measures proficiency in speaking.

Instruction in the performance objectives was introduced in 1987 and test batteries for 13 languages were developed and fully implemented by 1994. For each language, there is a series of ten performance objectives test. These tests are task-oriented, constructed-response tests, as opposed to multiple-choice tests. For example, examinees are asked to produce an English summary of a conversation, transcribe text in the target language, read legible native handwriting, translate transcribed materials, etc.

DLIFLC has three major types of students: cryptologists, human intelligence personnel and Foreign Area Officers. Approximately 70 percent of the students are cryptologists. The majority of cryptology students attend a follow-on school at Goodfellow Air Force Base (GAFB) in San Angelo, Texas, where they receive job-specific training involving foreign language skills. The cryptology students attending GAFB are drawn from all four uniformed services. Of the twenty-one languages taught at DLIFLC, the ten highest enrollment languages have a follow-on component at GAFB. Graduates of the other twelve languages, which account for approximately 30 percent of the enrollees, do not go to follow-on training at GAFB. Because job requirements can vary for the different services, in some languages GAFB offers different courses for members of the different services. Each GAFB course consists of a series of "blocks" of instruction reflecting training objectives for that course. GAFB evaluates its training within these courses with tests based on these blocks, some of which are multiple choice and some of which are of the constructed-response type.

B. PROBLEM

In the past, general proficiency in listening, reading, and speaking skills has been the focus of language learning and testing in the DoD. A general rule applicable for all services is that cryptology students with a minimum acceptable DLPT score (measuring general proficiency) are eligible to attend follow-on training at GAFB.

DoD does not have a corresponding rule establishing minimum acceptable performance objective scores for entry into GAFB. Individual service commanders at DLIFLC may exercise some discretion in borderline cases where general minimum DLPT requirements have not been met. They may take into account a variety of factors, such as motivation, military bearing and performance objective scores, to grant waivers for attending GAFB follow-on training.

Some GAFB "block" tests are similar to performance objectives tests in format and language skills addressed; for this reason the DLIFLC Evaluation Division believes the performance objectives test scores can be an extremely important factor in determining the probability of success in follow-on training and ultimately the field.

The purpose of this study is to determine how the performance objectives test scores relate to success on the DLPT and how combinations of DLPT and performance objective tests might relate to success in follow-on training at GAFB. For the purpose of this study, success at GAFB is defined by on-time graduation, number of required mandatory study hours, and performance on "block" tests.

The results of this study will assist Service Commanders in interpreting the meaning of performance objectives tests when making decisions about waivers for admission to GAFB follow-on training. The results may also be of interest to language departments and service commanders in making decisions about recycling students prior to graduation. Recycling means returning a borderline student to an earlier point in the course in a trailing class in order to give the student time to work on academic weaknesses. The results of this study might also help interpret the meaning of tests given prior to graduation that are similar to either the DLPTs or performance objectives in format and content.

C. ORGANIZATION OF THESIS

Chapter II contains a review of the literature on prediction of success at DLIFLC. Chapter III describes the data and variables considered. Chapter IV outlines a description of the method used to analyze the data. Chapter V contains the findings of the analysis. Chapter VI contains a discussion on the summary, conclusions and recommendations. The statistical package used in this thesis is named SPSS (Ref. 10). The Appendices present an example of the SPSS output, graphs that show predicted values on tests to achieve a predetermined probability of passing designated DLPTs or "block" tests, and the S-plus code used to create the graphs.

II. LITERATURE REVIEW

While there is a large literature on the learning of language in civilian schools, the military has gone largely un-analyzed. The issue of predicting language learning success has been analyzed in a few other studies. However, a formal study dedicated to correlation of performance objectives and proficiency FLOs with follow-on training measures has not been performed, nor has a formal study been conducted on the correlation of performance objectives and proficiency FLOs within each language. The following are brief descriptions of the previous research conducted on predicting language learning success completed at DLIFLC.

A. LANGUAGE SKILL CHANGE PROJECT

The Army Research Institute for the Behavioral and Social Sciences and the DLIFLC conducted a joint research effort to determine the effectiveness and efficiency with which foreign language skills are learned, retained, and applied to job responsibilities in the Army. The specific objectives of the study were to 1) track changes in language proficiency over time, 2) identify factors related to changes in proficiency, and 3) better understand predictors of language learning at DLIFLC. The Language Skill Change Project (LSCP) (Ref.3) was a longitudinal study that followed approximately 2000 Army linguists throughout their foreign language training and in their first tour of duty in the field. Data were collected from the linguists at seven different times starting from the first week of their language training at DLIFLC and extending until approximately three years after their graduation from DLIFLC.

Report II of LSCP, entitled "The Prediction of Language Learning Success at DLIFLC," (Ref. 6) indicated that success can be predicted by non-cognitive measures. The findings support the continuation and expansion of linguist select procedures based on cognitive ability for admission to DLIFLC training. Of all the types of student characteristics considered in this research, the measures of the different cognitive aptitudes had the greatest success as predictors of performance. In developing improved selection procedures, however, some consideration should be given to the possibility of incorporating at least some non-cognitive attributes as well. Specifically student attitudes, motivation and applied learning strategies made significant contributions to the prediction of listening and reading skills. Motivation, provided relatively important prediction increments to the less predictable speaking skill. Report III of LSCP, "Training Approaches for Reducing Student Attrition From Foreign Language Training," (Ref. 5) showed that in the samples studied, a Defense Language Aptitude Battery (DLAB) score of 100 was pivotal in determining trends for attrition. Students with scores of 100 or below were more likely to attrit than those students with scores above 100.

B. OTHER DLIFLC RESEARCH

1. "Language Choice and Performance."

The Research and Analysis Division (ESR) of the DLIFLC was tasked to investigate whether the level of proficiency attained by students in the Basic course has a relationship to whether or not the language assigned was their language of choice. The study (Ref. 4) was conducted on a sample of Fiscal Year (FY) 1990-1994 graduates of the DLIFLC Basic course in eight languages. This study indicated that there was minimal correlation between ability to choose which language to study and subsequent performance in the language studied; thus, other factors should be chosen to explain training outcomes.

2. "The Effects of Length of Service and Prior Language Study at DLI on DLPT Attainment."

This study (Ref. 7) was conducted by ESR to compare the DLPT performance of enlisted military personnel who had four or more years of service to that of initial entry trainees (IET), who had less than one year of service before enrolling in DLIFLC Basic Language Course. Additionally, the study covered those who had studied a language at DLIFLC prior to their current enrollment to those who had not. This study showed no significant difference in performance between IETs and those personnel with more than four years of service. The results do, however, strongly support the use of previous foreign language study as a useful predictor of subsequent language learning success. Aptitude measures had statistically significant correlation with proficiency in all three skills.

3. "Relationships of Language Aptitude and Age to DLPT Results among Senior Officer Students in DLIFLC Basic Language Courses."

ESR conducted this study (Ref. 8) pursuant to the request from the DLIFLC Command Group to examine the relationships of age and aptitude among all basic course students in paygrades O5 and O6. The results were that correlation of age with DLPT measures of listening, reading and speaking were not statistically significant.

III. THE DATA

Personal and career statistics of students who have attended DLIFLC and GAFB are maintained in a database at DLIFLC. The data for this study were obtained from this database.

A. THE POPULATION

The majority of the training at DLIFLC is conducted in the basic acquisition courses of language instruction. The Basic course is largely composed of enlisted military students who have one or fewer years of military service.

In the first phase of the study, we examine the relationships between performance objectives in various languages and proficiency DLPTs for all students graduating from DLIFLC between the beginning of FY96 and the end of FY97. This data set includes records for 5413 students.

In the second part of the study, we consider both proficiency and performance FLOs as predictors of measures of success in follow-on training at GAFB for a subsample of the original population. This subsample includes only students of Russian. The dependent variables in this subsample were different for students in each Service, because GAFB has different courses with different criterion measures for the Army, Air Force, and Navy/Marine services. This overall Russian subsample included 516 records.

B. THE VARIABLES

1. First Portion of Study: Dependent Variables

The dependent variables for the first portion of this study were the scores obtained on the DLPT. The DLPT is used as the standard for successful completion of the initial course of language instruction. There are three scores on the DLPT for each language: the first is for listening, the second is for reading, and the third is for speaking.

The DLPT speaking, listening, and reading scores are reported on a scale with eleven points; each point is called a "level score." Within the U.S. Government and DoD, speaking, listening, and reading scores are reported on a scale with eleven possible levels. The possible level scores are 0, 0+, 1, 1+, 2, 2+, 3, 3+, 4, 4+, and 5. Levels 3+, 4, 4+, and 5 in listening and reading are not awarded at DLIFLC for reading and listening, however the full range of score may be awarded for DLPT in speaking. The scale of level scores indicates levels of proficiency for military linguists as defined by verbal descriptions approved by the Federal Interagency Language Roundtable. There is a general rule applicable to all Services that students with at least Level 2 in Listening, Level 2 in Reading, and Level 1 in Speaking are eligible to attend follow-on training at GAFB. Level 2 in reading is described as sufficient comprehension to read simple, authentic written material in a form equivalent to usual printing or typescript on subjects within a familiar context. A Level 2 student will therefore be able to read texts that are normally presented outside of a classroom environment, for example a newspaper clipping or business letter. A Level 2 listening score is defined as sufficient comprehension to understand conversations on routine social demands and limited job requirements (e.g., be able to understand face-to-face speech in a standard dialect, delivered at a normal rate by a native speaker not used to dealing with foreigners, about everyday topics). The speaking Level 1 is defined as the ability to satisfy minimum courtesy requirements and maintain very simple face-to-face conversations on familiar topics. For example, this speaker would be able to ask for help and verify comprehension of a native speaker, but misunderstandings would be frequent. The DLPT speaking score is obtained directly from an interview conducted by trained and certified language testers. The DLPT in listening and reading yield converted scores of 0 to 60, which yield level scores ranging from 0 to 3. For this analysis, the converted scores was used for the reading and listening tests.

2. First Portion of Study: Independent variables

The independent variables used in the first portion of the study for each language sample were the ten performance objectives test scores. (While other variables might have been considered, the objective as stated by DLIFLC-ESR is to make predictions based on scores on these tests.) The possible scores on each of the ten performance objectives range from 0 to 100. Table 1 is a description of the performance objective categories:

Table 1. Description of Performance Objective Categories

TEST NUMBER	CATEGORY	SKILL
F1A	Listening	Produce an English summary of a news broadcast or conversation.
F2A	Listening	Answer content questions about a news broadcast or conversation.
F3A	Transcribing	Transcribe text into native script.
F4A	Transcribing	Transcribe decontextualized numbers.
F5A	Reading	Answer content questions about a level 2 written text.
F6A	Reading	Read reasonably legible hand- written native text.
F7A	Translation	Translate level 2 text into idiomatic English.
F8A	Translation	Translate an English text into level 2 target language.
F9A	Speaking	Biographical data interview.
F10A	Speaking	Two-way interpretation.

3. Second Portion of Study: Dependent Variables

The dependent variables for the second portion of this study were the scores obtained on the GAFB "block" tests in the respective Russian courses for the various services, the total time to train at GAFB, and the number of hours required in the Special Individual Assistance (SIA) program.

The "block" test scores are obtained from a variety of different tests. Some of these tests yield a pass/fail score while others have a score that ranges from zero to one hundred.

Special Individual Assistance is a program developed for those students who are having difficulty in the course of instruction. GAFB mandates special hours of additional help in the areas in which these students are having difficulty.

4. Second Portion of Study: Independent Variables

The independent variables for the second portion of the study include both performance objectives and DLPTs as discussed above, but only for the Russian subsample.

IV. METHODOLOGY

A. REGRESSION ANALYSIS MODEL

Regression analysis models allow the forecaster to estimate the value of one variable based on its relationship to one or more other variables. Simple regression assumes that the functional relationship between two variables can be represented as a straight line. Each of the n observations is assumed to obey:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \; i = 1, \ldots, n \tag{1}$$

where Y_i is the i^{th} value of the dependent variable, X_i denotes the corresponding value of the independent variable, β_0 is the point at which the straight line intersects the Y-axis, β_1 is the regression coefficient or slope of the line, and ε_i is the "error" which describes the departure of this observation from the line. Simple regression uses the ordinary least squares (OLS) method to find the equation for a straight line which most closely approximates the underlying data set (Ref. 2, pp. 30-33). Multiple regression is identical to the simple regression model except that the model uses multiple (say, k–1) predictors (X's) for each data point. The least squares method then fits a plane rather than a straight line:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots\ldots + \beta_k X_{ik-1} + \varepsilon_i, \; i = 1, \ldots, n \tag{2a}$$

or, in matrix notation,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2b}$$

where $\mathbf{Y}$ is an n-vector of observations of the dependent variable; X ($n \times k$) is the matrix of observations of independent variables (here including a column of 1's for the intercept), $\boldsymbol{\beta}$ is the k-vector of regression coefficients (here including the intercept, β_0), and $\boldsymbol{\varepsilon}$ is the n-vector of "errors." (Ref 2, p. 66).

B. THE GENERAL MODEL

Ordinary least squares multiple linear regression analysis is used to fit the model of each dependent variable to the data available. The least-squares principle specifies that the b_j's (estimated coefficients) are to be chosen so as to minimize the sum of squared differences between the observed values and the estimated values of the dependent variable. This quantity is known as the sum of squared residuals (RSS).

$$RSS = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \tag{3}$$

or in matrix terms:

$$RSS = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) \tag{4}$$

where the superscript "T" denotes transposition.

We estimate the vector $\boldsymbol{\beta}$ (the true coefficients), by the solution, $\mathbf{b}$, to the following equation (Ref 2, p. 72):

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y}) \tag{5}$$

C. THE STEPWISE REGRESSION MODEL

Stepwise regression is an automatic method of building a multiple linear regression model to select the set of independent variables for inclusion. This procedure can be described as a step-up procedure with a step-down adjustment. First, starting with no X variables in the model, the computer program chooses the variable that has the largest simple correlation with Y. Thereafter, it either adds the X variable that produces the largest further increase in R^2 or removes the variable that will least reduce R^2 (see section D). At each step the p-value for the usual F-test is computed. The procedure stops when a specified significance level, .05 for forward selection and .1 for backward elimination, cannot be met by any further inclusion or exclusion of a variable (Ref 1, p. 123). This selection procedure does not guarantee optimum subsets, but it does overcome some of the major deficiencies encountered in other methods and is the best method offered by SPSS.

D. THE R^2 STATISTIC

A commonly accepted statistic for measuring the value of a regression equation is the R^2 statistic. The R^2 statistic measures the proportion of total variation about the mean which is accounted for by the regression, equation (6). This statistic should be viewed with some caution, because it can be made arbitrarily high by adding additional variables; nonetheless it is widely used and so we report it here.

$$R^2 = \frac{ExplainedVariance}{TotalVariance} = \frac{ESS}{TSS} = \frac{\sum(\hat{Y}_i - \overline{Y})^2}{\sum(Y_i - \overline{Y})^2} \tag{6}$$

where $\hat{Y}_i$ is the i^{th} predicted value, $\overline{Y}$ is the mean of the dependent variable, and Y_i is the i^{th} actual value (Ref 2, p. 39).

E. THE t-TESTS AND F-TEST

The OLS yields estimates (b_j) for our regression coefficients β_j. The estimated standard error, $\hat{\sigma}$, of the regression is

$$\hat{\sigma} = \sqrt{\frac{\sum(Y_i - \hat{Y}_i)^2}{n-k}} \tag{7}$$

where k is the number of estimated parameters and n is the number of data points (Ref 2, p. 36).

Assuming that errors are independently and identically distributed as $N(0,\sigma^2)$, the statistic

$$t_j = \frac{b_j - \beta_j}{SE_{b_j}} \tag{8}$$

where SE_{b_j}, the standard error for the estimated coefficent b_j, is the j^{th} diagonal element of the estimated covariance matrix of the parameters,

$$\hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1}, \tag{9}$$

follows a Student's t-distribution with $n-k$ degrees of freedom (where n is the number of data points) under the null hypothesis

$$H_0: \ \beta_j = 0. \tag{10}$$

The p-value is the estimated probability of obtaining results as extreme as the sample or more extreme when the data is drawn from a population in which H_0 is true. A low p-value indicates that it is unlikely that such a sample would come from a population where H_0 is true; therefore we can reject the null hypothesis and state that it is likely that there is a linear relationship between the dependent and independent variables. The critical value used to reject the null hypothesis in this study is .05. Therefore any p-value obtained less than .05 is said to be "statistically significant."

An F-statistic can test hypotheses regarding sets of parameters. The null hypothesis for this test is,

$$H_0: \ \beta_1 = \beta_2 = \ldots\ldots = \beta_k = 0 \tag{11}$$

Using the same philosophy as with the *t*-statistic, we reject the null hypothesis if the *p*-value is less than .05:

$$F = \frac{\dfrac{\sum(\hat{Y}_i - \bar{Y})^2}{df_R}}{\dfrac{\sum(Y_i - \hat{Y})^2}{df_E}} \tag{12}$$

where df_R and df_E are the degrees of freedom for the regression and the error respectively (Ref 2, pp.43-45). We will discuss the assumptions and limitations that we used in the model in Sections F and G.

The use of a regression model to analyze a set of data is subject to a number of assumptions and limitations (Ref. 2, pp. 110-112).

F. ASSUMPTIONS

1. Fixed X

In this study, the X values are not fixed as part of the design. Therefore we proceed with the anlysis conditional on the X's we actually observe.

2. Errors are normally distributed with a mean of zero

This means that over the long run, sample estimates (b_k) will center on the true parameter value (β_k). A probability plot and histogram of residuals are observed to verify that errors are Normally distributed. These plots are produced as a matter of course by the SPSS software; see the example in Appendix A. In general, the assumption of Normality seems to be approximately correct. The assumption that the mean of the errors is zero cannot be tested, since the residuals always have mean 0; however, the consequences of a non-zero mean are limited to a bias in the intercept (β_0) term.

3. Homoscedasticity (errors have constant variance)

The third assumption is that the variance of the regression errors is constant. The variance of these errors, also known as residuals, must remain constant over the entire range of values for the independent variable. Variables with non-constant variances can give significance tests that are meaningless. To verify that homoscedasticity exists, thereby validating the assumption of constant variance, a residual versus predicted values plot is observed. (See Appendix A for an example.) The plot should show a random pattern, and this assumption generally appears valid.

4. Errors are uncorrelated with each other (no autocorrelation)

The fourth assumption we used is that the errors are independent of one another. This assumption should be safe because the observations are not collected at points adjacent in time or space. Interestingly, the usual Durbin-Watson test showed occasional departures from this assumption, but given the nature of the data it is difficult to explain serial correlation. We proceed as if this assumption were correct.

G. LIMITATIONS

1. Omitted Variables

If other variables affect both X and Y, b_j may substantially overstate or understate the true relationship between X and Y. Of course we cannot identify these variables.

2. Nonconstant Variance of Errors (Heteroscedasticity)

If the variance of the errors were to vary with the level of X, the usual standard errors, hypothesis tests, and confidence intervals would not be trustworthy. In small samples it can be difficult to assess the residual versus predicted plots. The assumption of homoscedasticity does seem to hold in the large-sample cases.

3. Nonlinear Relationships

OLS finds the best-fitting straight line. This can be misleading if the expected value of Y_i is a nonlinear function of X. A pattern in the plot of residuals versus fitted values (see Appendix A) would be evidence of a violation of this assumption, but such a pattern was not seen.

4. Non-Normal Errors

The usual *t* and *F* procedures assume that the residuals are Normally distributed. This assumption seemed to hold up in the large-sample cases and is difficult to assess in smaller samples. When errors are non-Normal, p-values from these procedures are untrustworthy.

5. Influential Cases

OLS can be affected by outliers, which can pull the line up or down and substantially influence all results. This was examined primarily in the unusual cases where performance objective coefficients were negative. There was no evidence of recording errors in the data.

H. DETERMINATION OF PROBABILITY SCORES

1. Single Main Effect Models

The determination of performance objectives cut-off scores for the models with one main effect was conducted using the assumption that the errors in the model are Normal, thereby ensuring that the predicted DLPT scores are also Normal. For any specific (row) vector of independent variables $\mathbf{X}_0$, the model predicts the value $\hat{Y}_0 = \mathbf{X}_0\mathbf{b}$.

The standard error of this prediction is given by $\mathrm{SE}(\hat{Y}_0) = \hat{\sigma}\,(1 + \mathbf{X}_0\,(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}_0^{\mathrm{T}})^{1/2}$ (Ref.2, p. 79).

The distribution of the DLPT for a specific performance objective score is then:

$$\mathrm{N}(\mathbf{X}_0\mathbf{b}\,,\ \mathrm{SE}(\hat{Y}_0)^2\,) \tag{13}$$

and the quantity $(Y_0 - \mathbf{X}_0\mathbf{b}\,)/\mathrm{SE}(\hat{Y}_0)$ should follow the Standard Normal. We seek the performance objective score for which we predict an 80% chance of reaching a pre-determined cut-off (a passing score) on the DLPT. Thus we have

$$\Pr\left(\frac{Y_0 - \mathbf{X}_0\mathbf{b}}{\mathrm{SE}(\hat{Y}_0)} > \frac{\text{cut-off} - \mathbf{X}_0\mathbf{b}}{\mathrm{SE}(\hat{Y}_0)}\right) = 1 - \Phi\left(\frac{\text{cut-off} - \mathbf{X}_0\mathbf{b}}{\mathrm{SE}(\hat{Y}_0)}\right) = 0.8 \tag{14}$$

from which we get

$$\mathbf{X}_0\mathbf{b} = \text{cut-off} - \mathrm{SE}(\hat{Y}_0) \times \Phi^{-1}(.2). \tag{15}$$

In a model with only one independent variable, we can then find the performance objective score for which the predicted probability of a passing score (40 for DLPT_R or DLPT_L, 20 for DLPT_S) is 80%. In fact, we draw a graph of performance objective score (X) against predicted probability of passing for every X, for every model with only one main effect, and plot them in Appendices B (for DLIFLC) and C (for GAFB). These plots were constructed by the software package S-Plus (Ref. 9)

For example on the DLPT_L for Czech, see Figure (1). A score of 93 or greater on performance objective 9 needs to be obtained to have an 80 percent chance of reaching a score of 40 or greater on the DLPT_L.

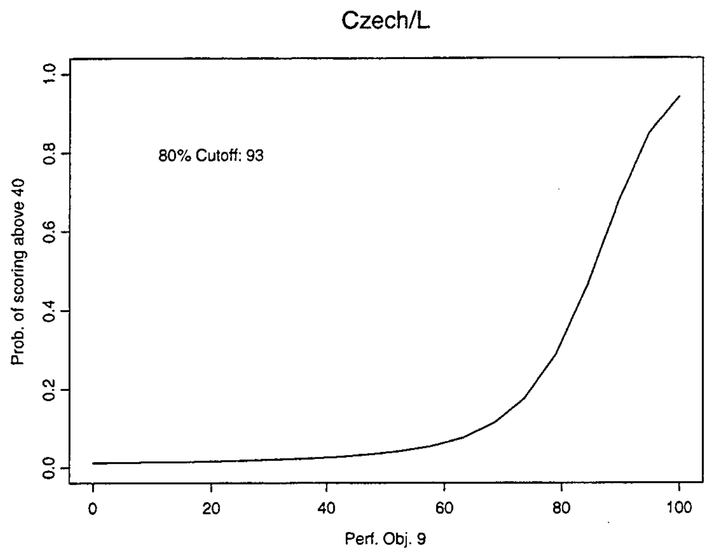


Figure 1. Probability of Scoring 40 or Greater on DLPT_L/Czech Given F9A

This graph has the sort of shape we expect: a student who scores poorly on test F9A is predicted to have little chance of passing the DLPT_L and a student who does very well is predicted to have a high chance of passing. Some of these graphs have a less intuitively-appealing shape, however. For example, it appears that most students pass the DLPT_L in Tagalog, regardless of their scores on the "best" predictor, test F1A. On the other hand, even a student who scores very well on test F7A does not have a predicted probability of 80% of passing the Korean DLPT_S. See Appendix B.

2. Models with Two Main Effects

A similar analysis can be done in a model with two main effects. In this case, X_0 contains two performance objectives and there will be an infinite number of combinations of scores for which the predicted DLPT score yields an 80% chance of passing. We can plot the "frontier" made up of all such combinations for every model with two main effects. The S-plus code for developing this frontier graph is in Appendix E.

For example as shown in Figure (2), for the GAFB Navy Russian students, some of the combinations of scores on F8A and F7A for which the predicted probability of scoring a 70 or greater on "block" test 27 is 80% are:

F8A = 60; and F7A = 7;

F8A = 40; and F7A = 16; and

F8A = 5; and F7A = 35.

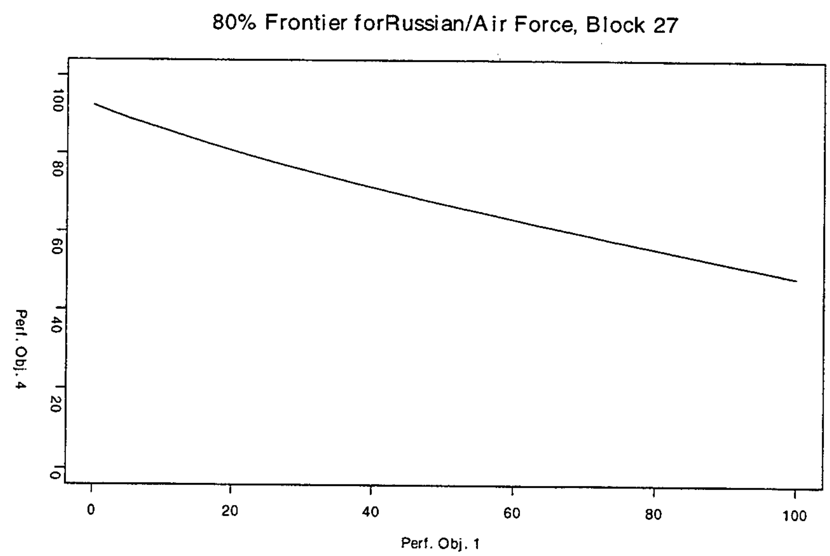


Figure 2. Eighty Percent Probability of Scoring 70 on Block Test 27/Navy Given F8A and F7A

Of course, any combination of scores whose position on the graph is above and to the right of the line leads to a predicted probability greater than 80%. Some interesting features can be seen on these graphs (see Appendices E and F). For example, in a number of the DLPT cases the frontier is very near the right-hand corner of the graph, showing that very few combinations of scores yield a predicted probability of passing as high as 80%. Conversely, at GAFB it is often the case that every student passes, so that the frontier coincides with the co-ordinate axes. (In those cases no picture is supplied.) As discussed in section V.A, it sometimes happens that the regression coefficients are negative. The effect of this on the frontier graph can be seen in, for example, Czech on the DLPT_R. The frontier has a positive slope, indicating that students with higher scores on F1A need *higher* scores on F5A to reach a predicted 80% probability of passing the DLPT_R than students with lower scores on F1A. This result is clearly counter-intuitive.

V. FINDINGS

The first criterion for selecting a model was that the F-statistic comparing the null model to the model with a single term be significant at the 5% level, which indicates that the model is better than simply using the mean of the dependent variable. Originally we considered models with interactions and models growing out of factor analysis. However, DLIFLC found these to be un-interpretable. Furthermore the decreases in standard error obtained with these models, compared to models with only main effects were minimal. Thus every model had only main effects for the independent variables.

Once a single-term model had been chosen, our second criterion came into play. That was that in our judgment, a decrease in standard error of less than 0.1 did not warrant the addition of another term to the model, even if that term was "statistically significant" by the regression *F*-test. Such a term was deemed to be of no practical significance. Starting with a one-term model, then, terms were added one at a time until adding a term caused an improvement in standard error less than 0.1. For example, SPSS produced seven possible models (all with significant *F*-statistics) for the Arabic DLPT_L model, one each with one main effect, two main effects, and so on up to seven main effects. The model with one main effect had a standard error of 4.17, the model with two main effects 3.74, the model with three main effects 3.57, and the model with four main effects 3.51. Since the difference in standard error for the model with one main effect (4.17– 3.74 = 0.43) was greater than 0.1, we then considered the model with two main effects. A similar subtraction comparing the standard errors for models of size two and three (3.74 – 3.57 = 0.17) also gave a result greater than 0.1. For the third model, the difference (3.57

– 3.51 = 0.06) was less than 0.1; therefore this model was chosen. The order of the variables within the models is the order in which the stepwise regression entered the variables. For example, in the Arabic model for DLPT_L, F2A was the first variable to enter the model, then F7A, and lastly F1A. See output in Appendix A.

A. LISTENING PROFICIENCY MODELS BY LANGUAGE

A summary of the DLPT_L models by language are shown in Table 2. In that table, "STD ERR" denotes the standard error of the regression, while "STD DEV" gives the standard deviation of the responses. (This, of course, is also the standard error from the naïve model that includes only an intercept.) "ABC S/D" indicates whether alphabets are similar to our own (that is, Roman) or different from it; "LAN CAT" gives the language's category of difficulty.

Originally we thought that the performance objectives scores that would measure the listening proficiency were F1A, F2A, and possibly F3A and F4A. The variable which occurred most frequently across all languages for the DLPT_L were F1A, F3A, and F7A as shown in Table 3. These variables appear to be best predictors of performance on the DLPT_L. Furthermore, among languages using similar alphabets (the Roman alphabet), F1A and F3A appeared to be the best indicators, while for dissimilar alphabets F1A and F7A appeared to be better indicators as shown in Table 5. Additionally, F3A and F8A appeared to be the best predictors in Category I languages. There is only one language in Category II, so there was no analysis done for this category. In Category III languages, F1A and F7A appeared to be the best predictors, and in Category IV languages F1A and F3A were the best predictors. See Table 4. It is interesting to note that in some cases, one or more of the regression coefficients is negative. This indicates (counter to our expectations) that an increase in the performance objective score is associated with a decrease in the predicted DLPT score. The reason for this result may be that there really are interactions between independent variables that our model does not include. Our standard errors of prediction are generally about as small as in models that include interactions, however. (See also section V.E.2.) Additionally, we note that while the performance objective scores are all on the same scale, the estimated coefficients can vary by a factor of about one-thousand (ranging from about 0.4 to about 0.0008). In each case, though, the addition of a term reduces the standard error of the regression model by at least 0.1.

Table 2. Summary of Models for DLPT_L

LANGUAGE	EQUATION	STD ERR	N	R^2	STD DEV	ABC S/D	LAN CAT
Arabic "A"	$27.708 + .116 * F2A + 8.754*10^{-2} *F7A + .477*10^{-2} *F1A$	3.57	712	.568	5.42	D	IV
Chinese-Mandarin "C"	$38.564 + 7.601*10^{-2} * F6A + 6.597*10^{-2} * F1A$	2.63	218	.396	3.36	D	IV
Czech "Z"	$.945 + .462 * F9A$	3.87	19	.335	4.71	S	III
French "F"	$4.755 + .288 * F3A + .193 *F8A$	3.99	121	.449	5.34	S	I
Hebrew "H"	$22.940 + .265 * F7A$	3.97	52	.518	5.81	D	III
Japanese "J"	$-.242 + .458 * F5A + .179 * F4A + .156 * F7A - .321 * F1A$	2.44	23	.784	5.11	D	IV
Italian "I"	$11.898 + .363 * F8A$	4.50	43	.442	5.73	S	I
Korean "K"	$33.405 + .102 * F1A + 5.281*10^{-2} * F3A$	2.78	427	.342	3.46	D	IV
Persian-Farsi "P"	$34.492 + .109 * F2A + 6.024*10^{-2} * F5A + 5.081*10^{-2} * F3A$	3.05	223	.484	4.27	D	III
Polish "L"	$-18.950 + .232 * F1A + .719 * F10A$	2.95	13	.771	5.43	S	III
Spanish "S"	$30.549 + .131 * F1A + 5.371*10^{-2} * F3A + 7.304*10^{-2} * F7A$	3.98	778	.417	5.30	S	I
Russian "R"	$36.76 + .126 * F1A + .113 * F2A$	3.52	594	.476	4.91	D	III
Tagalog "G"	$41.530 + 7.945*10^{-2} * F1A$	2.10	17	.359	2.54	S	III
Thai "T"	$6.782 + .371 * F4A$	6.45	31	.455	9.88	D	III
Vietnamese "V"	$19.902 + .221 * F9A + 9.917*10^{-2} * F1A$	3.86	74	.248	4.39	D	III

Table 3. Frequency of Variables in the DLPT_L Model

VARIABLES	FREQUENCY	PERCENT
F1A	9	60
F2A	3	20
F3A	4	27
F4A	2	13
F5A	2	13
F6A	1	7
F7A	4	27
F8A	2	13
F9A	2	13
F10A	1	7

Table 4. By Category of Difficulty

	Category 1 (3)		Category 3 (8)		Category 4 (4)	
Variables	Frequency	Percent	Frequency	Percent	Frequency	Percent
F1A	1	33	4	50	4	100
F2A	-	-	2	25	1	25
F3A	2	67	1	12.5	1	25
F4A	-	-	1	12.5	1	25
F5A	-	-	1	12.5	1	25
F6A	-	-	-	-	1	25
F7A	1	33	1	12.5	2	50
F8A	2	67	-	-	-	-
F9A	-	-	2	25	-	-
F10A	-	-	1	12.5	-	-

Table 5. By Category of Alphabet

	SIMILAR ALPHABET (6)		DISSIMILAR ALPHABET (9)	
Variables	Frequency	Percent	Frequency	Percent
F1A	3	50	6	67
F2A	1	17	2	22
F3A	3	50	1	11
F4 A	-	-	2	22
F5A	1	17	1	11
F6A	-	-	1	11
F7A	1	17	3	33
F8A	2	33	-	-
F9A	1	17	1	11
F10A	1	17	-	-

B. READING PROFICIENCY MODELS BY LANGUAGE

Originally we thought that the performance objectives scores that would measure the reading proficiency were F5A, F6A, and possibly F7A and F8A. The summary of the DLPT_R models by language is show in Table 6.

The variable which occurred most frequently for the DLPT_R were F2A, F5A, and F7A as shown in Table 7. These variables appear to be the best predictors of performance. Among the languages using similar alphabets (the Roman alphabet), F8A appeared to be the best indicator, while for dissimilar alphabets F2A, F5A and F7A appeared to be better indicators as shown in Table 9. Additionally, F8A appeared to be the best for Category I languages. In Category III languages, F2A, F5A and F7A appeared to the best predictors, and in Category IV languages F5A and F7A were the best predictors. See Table 8.

Table 6. Summary of Models for DLPT_R

LANGUAGE	EQUATION	STD ERR	R^2	N	STD DEV	ABC S/D	LAN CAT
Arabic "A"	$35.689 + .101 * F7A + 7.817*10^{-2} * F5A$	3.17	.496	713	4.52	D	IV
Chinese-Mandarin "C"	$36.562 + .124 * F6A + 8.666*10^{-2} * F7A$	3.50	.522	221	5.17	D	IV
Czech "Z"	$41.239 + .120 * F5A - 5.130*10^{-2} * F1A$	1.30	.595	19	2.34	S	III
French "F"	$-5.585 + .248 * F10A + .200 * F3A + .166 * F8A$	4.31	.444	121	5.61	S	I
Hebrew "H"	$17.491 + .210 * F6A + .137 * F2A - .122 * F4A + .103 * F7A$	3.23	.569	52	4.73	D	III
Japanese "J"	$-16.160 + .207 * F8A + .318 * F5A + .144 * F7A$	3.09	.664	23	5.49	D	IV
Italian "I"	$23.680 + .282 * F8A$	3.48	.445	43	4.38	S	I
Korean "K"	$26.558 + .136 * F5A + 6.303*10^{-2} * F10A + 6.923*10^{-2} * F1A$	3.03	.486	434	4.22	D	IV
Persian-Farsi "P"	$33.812 + 9.762*10^{-2} * F7A + .102 * F2A + 7.049*10^{-2} * F6A$	4.64	.419	224	6.01	D	III
Polish "L"	$-33.449 + .357 * F3A + .580 * F10A$	2.99	.663	13	4.83	S	III
Spanish "S"	$37.917 + 8.466*10^{-4} * F2A + .110 * F5A$	3.64	.343	776	4.53	S	I
Russian "R"	$31.942 + 7.164*10^{-2} * F2A + 6.323*10^{-2} * F7A + 8.230*10^{-2} * F5A + 7.872*10^{-2} * F1A$	3.31	.526	582	4.93	D	III
Tagalog "G"	$34.599 + .169 * F8A + .123 * F2A - 9.843*10^{-2} * F4A$	1.66	.789	17	3.26	S	III
Thai "T"	$27.504 + .305 * F2A$	7.70	.327	30	12.53	D	III
Vietnamese "V"	$25.897 + .151 * F5A + .126 * F8A$	5.05	.308	75	5.99	D	III

Table 7. Frequency of Variables in the DLPT_R MODEL

VARIABLES	FREQUENCY	PERCENT
F1A	3	20
F2A	6	40
F3A	2	13
F4A	2	13
F5A	7	47
F6A	3	20
F7A	6	40
F8A	5	33
F10A	3	20

Table 8. By Category of Difficulty

	Category 1 (3)		Category 3 (8)		Category 4 (4)	
Variables	Frequency	Percent	Frequency	Percent	Frequency	Percent
F1A	1	33	1	12.5	1	25
F2A	1	33	5	62.5	-	-
F3A	1	33	1	12.5	-	-
F4A	-	-	2	25	-	-
F5A	1	33	3	37.5	3	75
F6A	-	-	2	25	1	25
F7A	-	-	3	37.5	3	75
F8A	2	67	2	25	1	25
F9A	-	-	-	-	-	-
F10A	1	33	1	12.5	1	25

Table 9. By Category of Alphabet

	SIMILAR ALPHABET (6)		DISSIMILAR ALPHABET (9)	
Variables	Frequency	Percent	Frequency	Percent
F1A	1	17	2	22
F2A	2	33	4	44
F3A	2	33	-	-
F4A	1	17	1	11
F5A	2	33	5	56
F6A	-	-	3	33
F7A	-	-	6	67
F8A	3	50	2	22
F9A	-	-	-	-
F10A	2	33	1	11

C. SPEAKING PROFICIENCY MODELS BY LANGUAGE

Originally we thought that the performance objectives scores that would measure the speaking proficiency were F9A and F10A. The DLPT_S models are summarized by language in Table 10. The variables, which occurred most frequently for the DLPT_S, were F1A, and F7A as shown in Table 11. These variables appear to the best predictors of performance. Among the languages using the similar alphabets (Roman alphabet), F1A and F8A appeared to be the best indicators, and for dissimilar alphabets F7A was the better indicator of performance as shown in Table 13. Additionally, Category I languages did not show a dominant performance objective as a predictor. In Category III languages, F1A was the best predictor, and in Category IV languages F1A, F7A and F10A were the best predictors. See Table 12 below.

Table 10. Summary of Models for DLPT_S

LANGUAGE	EQUATION	STD ERR	R^2	N	STD DEV	ABC S/D	LAN CAT
Arabic "A"	$7.333 + 9.536*10^{-2} * F1A + 8.269*10^{-2} * F10A$	3.49	.285	716	4.13	D	IV
Chinese-Mandarin "C"	$15.384 + 5.786*10^{-2} * F6A$	2.65	.144	223	3.00	D	IV
Czech "Z"	$-18.024 + .441 * F9A$	3.79	.323	19	4.41	S	III
French "F"	$-1.1019 + .146 * F3A + .101 * F8A$	2.97	.279	123	3.48	S	I
Hebrew "H"	$6.795 + .166 * F2A$	3.29	.269	52	3.83	D	IV
Japanese "J"	$11.867 + 9.324*10^{-2} * F7A$	2.46	.220	23	3.46	D	IV
Italian "I"	$-5.955 + .302 * F10A$	2.17	.500	43	2.97	D	IV
Korean "K"	$14.605 + 6.041*10^{-2} * F7A$	3.00	.140	431	3.26	D	III
Persian-Farsi "P"	$7.470 + .112 * F10A + 5.499*10^{-2} * F5A$	2.89	.263	228	3.37	S	III
Polish "L"	$12.270 + .235 * F1A$	3.87	.420	13	4.73	S	I
Spanish "S"	$18.119 + 7.460*10^{-2} * F1A$	3.06	.166	778	3.38	D	III
Russian "R"	$8.890 + 9.921*10^{-2} * F1A + 7.559*10^{-2} * F7A$	3.91	.295	590	4.18	S	III
Tagalog "G"	$5.297 + .180 * F8A$	3.03	.507	17	4.18	D	III
Thai "T"	$14.368 + 4.610*10^{-2} * F7A + 3.994*10^{-2} * F3A + 9.278*10^{-2} * F1A$	1.82	.727	28	3.37	D	III
Vietnamese "V"	$-8.547 + .261 * F9A + 9.219*10^{-2} * F8A$	3.51	.267	77	4.04	D	III

Table 11. Frequency of Variables in the DLPT_S Model

VARIABLES	FREQUENCY	PERCENT
F1A	5	33
F2A	1	7
F3A	2	13
F5A	1	7
F6A	1	7
F7A	4	27
F8A	3	20
F9A	2	13
F10A	3	20

Table 12. By Category of Difficulty

	Category 1 (3)		Category 3 (8)		Category 4 (4)	
Variables	Frequency	Percent	Frequency	Percent	Frequency	Percent
F1A	1	33	3	37.5	1	25
F2A	-	-	1	12.5	-	-
F3A	1	33	1	12.5	-	-
F4A	-	-	-	-	-	-
F5A	-	-	1	12.5	-	-
F6A	-	-	-	-	1	25
F7A	-	-	2	25	2	50
F8A	1	33	2	25	-	-
F9A	-	-	2	25	-	-
F10A	1	33	1	12.5	1	25

Table 13. By Category of Alphabet

	SIMILAR ALPHABET (6)		DISSIMILAR ALPHABET (9)	
Variables	Frequency	Percent	Frequency	Percent
F1A	2	33	3	33
F2A	-	-	1	11
F3A	2	33	1	11
F4A	-	-	-	-
F5A	-	-	1	11
F6A	-	-	1	11
F7A	-	-	4	44
F8A	2	33	1	11
F9A	1	17	1	11
F10A	-	-	3	33

D. PROBABILITY OF PASSING DLPT

The statistical software, S-plus, was used to determine cut-off scores for those models with one and two main effects. In models with more than two main effects, cut-off scores on the performance objective tests can only be shown in three (or more) dimensions. These pictures are difficult to show and interpret.

The cut-off scores were calculated by assuming Normal performance objective scores utilizing the model, and the standard error of the model. We calculated for each language the performance objective score for which we predicted an 80 percent probability of scoring a proficiency of level of two. A proficiency level of two is determined by a converted score of 40 or greater on the DLPT_L or DLPT_R, and a score of 20 or greater on the DLPT_S. The results are shown for models with one main effect in Table 14 for DLPT_L, Table 15 for DLPT_R, and Table 16 for DLPT_S. The results for models with two main effects are shown in Appendix F.

Table 14. Eighty Percent Chance of Scoring 40 or Greater on DLPT_L Given:

LANG	FXA	SCORE
CZECH	9	93
HEBREW	7	77
ITALIAN	8	88
TAGALOG	1	7
THAI	4	NA

Table 15. Eighty Percent Chance of Scoring 40 or Greater on DLPT_R Given:

LANG	FXA	SCORE
ITALIAN	8	88
THAI	2	63

Table 16. Eighty Percent Chance of Scoring 40 or Greater on DLPT_S Given:

LANG	FXA	SCORE
CHINESE-MANDARIN	6	NA
CZECH	9	94
HEBREW	2	97
JAPANESE	7	NA
ITALIAN	10	92
KOREAN	7	NA
POLISH	1	50
SPANISH	1	60
TAGALOG	8	97

E. QUALITY OF MODELS

1. R^2 as a Quality Indicator

Utilizing R^2 as an indicator of a "good" model, Figure 3, 4, and 5 show that some languages appear to produce better models than others. The letters in quotes of Tables 2, 6, and 10 represent the language. Figures 3, 4 and 5 indicate that Japanese ("J") and Polish ("L") have a high R^2 for both the DLPT_L and the DLPT_R models (R^2_R"J" = .664, R^2_L"J" = .784; R^2_R"L" = .663, R^2_L"L" =.771), but the R^2 is not very high in the DLPT_S (R^2_S"J" = .220, R^2_S"L" = .420). Tagalog ("G") has a high R^2 for DLPT_R model, and a moderate R^2 for the DLPT_S model, but not a very high R^2 for DLPT_L model (R^2_R"G" = .789, R^2_L"G" =.359, R^2_S"G" =.507). Additionally, Vietnamese ("V") has a low R^2 for all three proficiency tests; DLPT_S, DLPT_L and DLPT_R (R^2_R"G" =.308, R^2_L"G" = .248, R^2_S"G" =.267). It is not clear to us why different languages should have different predictability from their respective performance objective tests. This is something that DLIFLC ought to investigate.

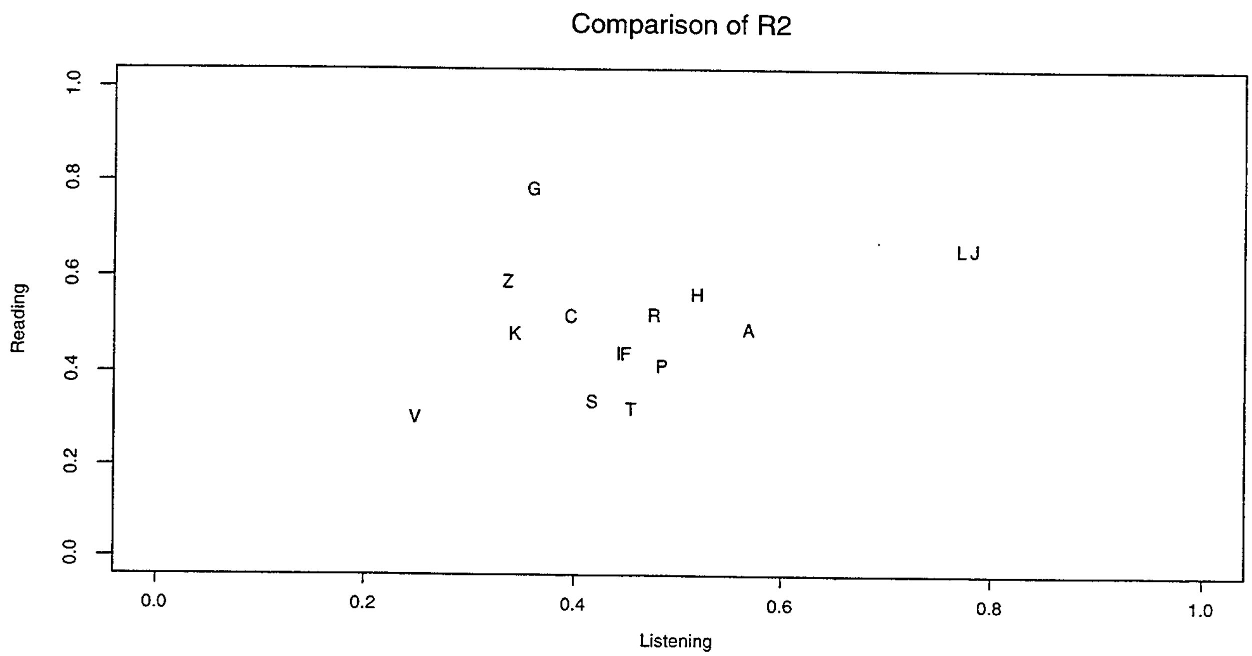


Figure 3. Comparison of R^2 For DLPT_L and DLPT_R.

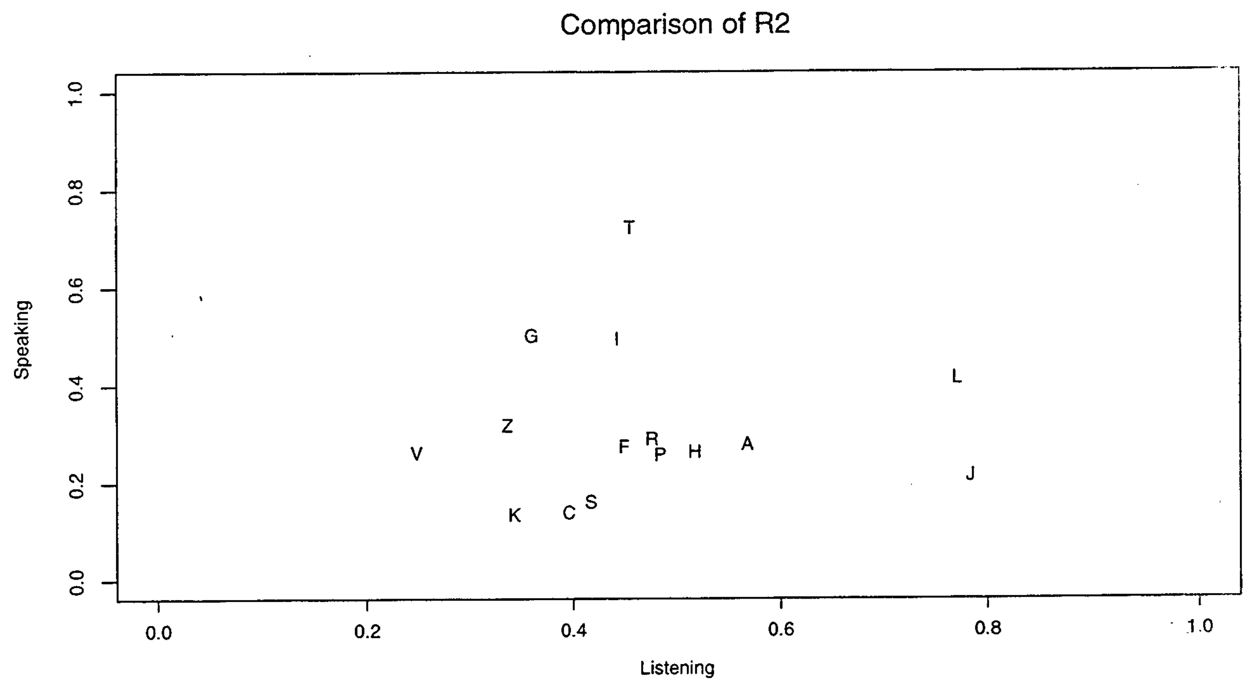


Figure 4. Comparison of R^2 for DLPT_L and DLPT_S.

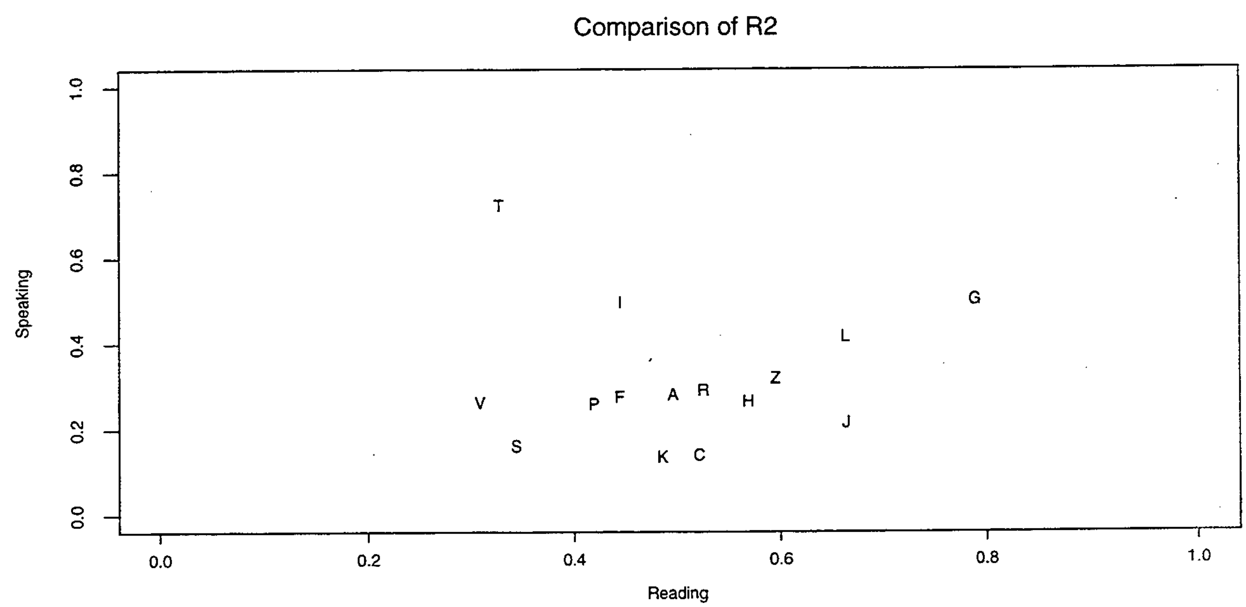


Figure 5. Comparison of R^2 for DLPT_R and DLPT_S

2. Models with Negative Main Effect Terms

In the Czech model, we saw that the results are very different from what we expected. For example, the relationship of performance objective 1 with the DLPT_R is "significantly" negative, when performance objective 5 is in the model. Possibly the performance objectives do not measure what they are supposed to, or we have seen a result of low probability. More likely, there is an interaction occurring amongst these variables, but when interactions are not allowed into the model, we get negative main effects. For example, Equation (19) would be the Czech model with interactions allowed (using the same criterion that allows in variables only if they reduce the standard error by more than 0.1):

$$DLPT\_R = 39.995 + 1.441*10^{-3} * F5F9 - 6.288*10^{-4} * F1F5 \qquad (19)$$

It is certainly reasonable for an interaction to be negative. One interpretation is that performance objective 1 is positively correlated with DLPT_R, and performance objective 5 is too, but performance objective 1 and performance objective 5 themselves have a highly positive correlation so that the effects when performance objective 1 and performance objective 5 are both high are not additive. Thus, someone who does "really" well on performance objective 1 and "really" well on performance objective 5 does better than someone who does well on performance objective 1 and well on performance objective 4, but the increase is not as much as one would expect.

F. GOODFELLOW AIR FORCE BASE FOLLOW-ON TRAINING

The same procedures that were used for DLPT models were used for the development of the GAFB models. One addition to the DLPT models was that the scores of the DLPT were considered as an independent variable for the prediction of success at the follow-on training at GAFB.

Each Service has a different course, each with different lengths and different tests. Therefore one cannot compare test scores for Army "Block 1" with Air Force "Block 1." A number of the "block" tests produced Pass/Fail grades.

The models in Tables 17, 19, and 21 were developed for those "block" tests which had variability in their scores, using SPSS. "NO MODEL" was placed in the "equation" column of the table for those "block" tests with no variability, such as for which every grade was "Pass." The data size for each course also varied, Army being the largest with 108 data points, Air Force with 35 data points and Navy/Marine Corps with 30 data points.

The most frequent variables for the Russian course were: for the Navy/Marine Corps F1A and F7A; for the Air Force F3A; and for the Army F5A, DLPT_S, and DLPT_R as shown in Tables 18, 20, and 22, respectively.

Table 17. Russian Navy/Marine Corps Models

BLOCK TEST	EQUATION	STD ERR	R^2	STD DEV
1	NO MODEL			
2	NO MODEL			
3	56.615 + .678 * DLPT_R	6.39	.159	7.09
4	NO MODEL			
5	NO MODEL			
6	NO MODEL			
7	62.838 + .193 * F3A	6.00	.248	6.83
8	80.438 + .499 * DLPT_S	3.09	.314	3.60
9	82.260 + .123 * F8A	5.22	.158	5.67
10	62.145 + .391 * F1A	6.78	.396	8.65
11	60.117 + .334 * F7A	4.65	.606	7.60
12	58.713 + .631 * DLPT_R	4.77	.227	5.34
13	68.613 + .201 * F7A	5.01	.323	6.57
14	74.063 + .194 * F5A	4.25	.225	4.77
15	92.921 + .200 * F7A – .524 * DLPT_S	3.80	.368	4.39
16	NO MODEL			
17	60.839 + .482 * DLPT_R	3.74	.218	4.58
18	89.142 + .205 * F1A	5.43	.219	5.80
19	82.854 + .434 * F1A – .721 * DLPT_S	5.12	.504	6.96
20	86.092 + .121 * F1A	3.25	.214	3.50
21	74.498 + .141 * F3A + $9.903*10^{-2}$ * F6A	3.61	.472	4.85
22	56.747 + .290 *F4A	4.98	.292	5.72
23	NO MODEL			
24	NO MODEL			
25	NO MODEL			
26	NO MODEL			
27	61.553 + .239 * F7A + .115 * F8A	4.50	.580	6.58
MSH	NO MODEL			
ACTL	NO MODEL			

Table 18. Frequency of Variables in the Goodfellow Russian Navy/Marine Corps Model

VARIABLES	FREQUENCY	PERCENT
F1A	4	24
F3 A	2	12
F4 A	1	6
F5A	1	6
F6A	1	6
F7A	4	24
F8A	2	12
DLPT_R	3	18
DLPT_S	3	18

Table 19. Russian Air Force Models

BLOCK TEST	EQUATION	STD ERR	R^2	STD DEV
1	NO MODEL			
2	NO MODEL			
3	NO MODEL			
4	24.779 + 1.262 * DLPT_L	11.18	.258	12.65
5	NO MODEL			
6	82.125 + .145 * F3A	4.40	.298	5.10
7	73.607 + .203 * F7A	7.08	.143	7.64
8	65.111 + .182 * F3A + .174 * F8A	4.49	.577	6.69
9	86.659 + .108 * F8A	4.27	.144	4.49
10	82.509 + .209 * F1A	6.43	.169	6.89
11	89.613 + .331 *DLPT_S	3.99	.130	4.18
12	88.125 + 9.788*10^{-2} * F3A	4.93	.134	5.20
13	NO MODEL			
14	66.879 + .312 F3A	7.23	.422	9.31
15	NO MODEL			
16	65.967 + .246 * F3A	8.16	.262	9.95
17	62.946 +.203 *F3A + .212 * F2A	5.98	.487	8.00
18	NO MODEL			
19	83.645 + .156 * F3A	5.27	.256	6.02
20	NO MODEL			
21	NO MODEL			
22	51.437 + 1.074 *DLPT_R – .387 * F5A – .968 * DLPT_S	6.53	.451	8.42
23	NO MODEL			
24	NO MODEL			
25	NO MODEL			
26	78.641 + .148 * F3A	8.05	.117	8.33
27	59.640 + .155*F1A + .262 * F4A	4.27	.365	5.13
MSH	NO MODEL			
ACTL	489.371 – 5.871 * DLPT_S + 5.719 * DLPT_R	39.39	.301	

Table 20. Frequency of Variables in the Goodfellow Russian Air Force Model

VARIABLES	FREQUENCY	PERCENT
F1A	2	12.5
F2A	1	6
F3A	8	50
F4A	1	6
F5A	1	6
F7A	1	6
F8A	1	6
DLPT_L	1	6
DLPT_S	3	19
DLPT_R	2	12.5

Table 21. Russian Army Models

BLOCK TEST	EQUATION	STD ERR	R^2	STD DEV
1	88.987 + .108 * F5A	4.41	.078	4.65
2	NO MODEL			
3	NO MODEL			
4	67.925 + .148 * F5A + .288 * DLPT_R	4.74	.223	5.22
5	31.971 + .393 * F4A + .683 * DLPT_S	7.18	.300	8.33
6	72.915 + .435 * DLPT_R	5.77	.096	6.26
7	NO MODEL			
8	NO MODEL			
9	75.234 + .657 * DLPT_S –.137 * F1A	8.66	.093	9.03
10	78.992 + .596 * DLPT_R – .151 * F10A	5.27	.196	5.79
11	NO MODEL			
12	NO MODEL			
13	20.085 + .619 * DLPT_S + .302 * F4A + .460 * DLPT_R	8.82	.229	10.32
14	75.844 + .194 * F7A	7.95	.106	8.85
15	88.355 + $8.861*10^{-2}$ * F5A	4.12	.061	4.17
16	NO MODEL			
17	NO MODEL			
18	40.114 + .233 * F5A + .590 * DLPT_S – .127 * F7A + .165 * F4A	5.67	.309	6.69
19	72.785 – .337 * F10A + .308 * F9A	4.04	.270	4.56
20	NO MODEL			
MSH	3.839 – .500 * F10A + .467 * F9A	4.61	.375	
ACTL	NO MODEL			

Table 22 Frequency of Variables in the Goodfellow Russian Army Model

VARIABLES	FREQUENCY	PERCENT
F1A	1	8
F4A	3	25
F5A	4	33
F7A	2	12.5
F9A	2	12.5
F10A	3	25
DLPT_R	4	33
DLPT_S	4	33

G. PROBABILITY OF PASSING "BLOCK" TESTS

As previously shown in the first part of the analysis for the DLPT, the statistical software, S-plus, was used to determine cut-off scores for those models with one and two main effects in the GAFB models.

The cut-off scores were calculated by assuming Normal performance objective scores and/or DLPT scores and utilizing the model and the estimate of the standard error of prediction, as before, to calculate the score required for an 80 percent probability of passing the "block" test. A passing score is 70 for the Navy/Marine Corps and Army, and 80 for the Air Force. The results are shown for models with one main effect in Table 23 for the Navy/Marine Corps, Table 24 for the Air Force, and Table 25 for the Army. (The corresponding graphs are shown in Appendix C.) The results for models with two main effects (in the form of "frontier graphs") are shown in Appendix G.

In a number of block tests, the grades were numeric (that is, not "Pass/Fail") and yet every student passed. That leads to scores of zero in tables 23-25. The implication is that regardless of the score on the performance objective, the probability that a student passes the "block" test is predicted as 100%. This explains graphs like the one for the Navy and Marine Corps Block 9, for example. In those graphs the "80%" level is reported as NA or 0 (the latter when a score only barely higher than 0 is required).

Table 23. Scores Required to Produce an 80% Chance of Scoring 70 or Greater on Russian Navy/Marine Corps Block Tests:

BLOCK	FXA/DLPT_X	SCORE
3	R	30
7	3	63
8	S	NA
9	8	NA
10	1	36
11	7	46
12	R	26
13	7	35
14	5	7
17	R	25
18	1	NA
20	1	NA
22	4	62

Table 24. Scores Required to Produce an 80% Chance of Scoring 80 or Greater on Russian Air Force Block Tests:

BLOCK	FXA/ DLPT_ X	SCORE
4	L	52
6	3	10
7	5	67
9	8	NA
10	1	15
11	5	NA
12	3	NA
14	3	62
16	3	86
19	3	0
26	3	53

Table 25. Scores Required to Produce an 80% Chance of Scoring 70 or Greater on Russian Army Block Tests:

BLOCK	FXA/DLPT_X	SCORE
1	5	NA
6	R	8
14	7	12
15	5	NA

VI. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

A. SUMMARY

In this thesis, we sought to accurately depict how the performance objectives correlate with the DLPT and how well the combination of the DLPTs and performance objectives forecast future performance at Goodfellow Air Force Base follow-on training. The models for each DLPT and each "block" test were created utilizing multiple linear regression method. In Chapter I, the background of the DLIFLC and the tests that are required for the analysis were discussed. In Chapter II, the previous studies were summarized. Chapter III gave a discussion of the population and the variables researched. In Chapter IV, the methodology for the model formulation was detailed. And finally, in Chapter V, the models and the statistics utilized to evaluate the accuracy of the models were summarized.

B. CONCLUSIONS

The primary research questions in this thesis are 1) What are accurate cut-off scores for the performance objectives and the DLPT to predict success at GAFB follow-on course? 2) How good are performance objectives for predicting future performance?

In some languages the performance objectives were better predictors of success on the DLPT than in others. For example, for the Polish language in the DLPT_L and DLPT_R, the R^2 statistic was high in both models and in the DLPT_S model, the R^2 was moderately high. Thus, to the extent that the R^2 statistic is an accurate indicator of a "good" model, then the performance objectives test are an accurate predictor for the DLPTs for the Polish language. However, the R^2 in the Vietnamese language was relatively low for all three DLPTs and therefore the performance objective test are not as accurate a predictor for the Vietnamese language.

Overall, performance objective 1, performance objective 3, and performance objective 7 were the most frequent performance objectives used as predictors for the DLPT_L. Performance objective 2, performance objective 5, and performance objective 7 were the most frequent performance objectives used as predictors for the DLPT_R. And finally, performance objective 1 was the most frequent performance objective test used as the predictor for success on the DLPT_S.

However, when divided by category of difficulty, performance objective 1 was the most frequent predictor of success on the DLPT_L for the more difficult languages. For the DLPT_R, performance objective 8 was the best predictor for Category 1 languages, performance objective 2 for Category 3 languages and performance objective 5 and performance objective 7 for the Category 4 languages. And finally, for the DLPT_S, performance objective 1 was the best predictor for the Category 3 languages and performance objective 7 for the Category 4 languages.

When the languages were divided by type of alphabet, performance objective 1 and performance objective 3 were the best predictors for the DLPT_L for similar (Roman) alphabets and performance objective 1 and performance objective 7 were the best predictors for the dissimilar (non-Roman) alphabets. For the DLPT_R, performance objective 8 was the best predictor for the similar alphabets and performance objective 2, performance objective 5, and performance objective 7 were the best predictors for dissimilar alphabets. For DLPT_S, performance objective 1 and performance objective 8 were the best predictors for similar alphabets and performance objective 1, performance objective 7 and performance objective 10 were the best predictors for the dissimilar alphabets.

In each language, different performance objectives were better predictors of DLPT tests scores. For example, F9A was the best predictor for DLPT_L in Czech, where F7A was the best predictor for DLPT_L in Hebrew. These performance objective tests were designed to measure proficiency in either Listening, Reading or Speaking. It appears that the performance objective tests are not measuring what they were intended for.

For the GAFB, again some of the proficiency tests were better predictors of success than others. For the Navy/Marine Corps Russian course, performance objective 1 and performance objective 7 were the best indicators of success for the "block" tests. Performance objective 3 was by far the best indicator for success for the "block" tests for the Air Force Russian course. And finally, performance objective 5, and DLPT_R and DLPT_S were the best indicators for success for the Army Russian course. Additionally, the proficiency tests at DLIFLC were not good indicators for predicting the number of mandatory study hours ("MSH") and the actual course length ("ACTL") for the GAFB Russian courses. Lack of variability in the course length and number of mandatory study hours in the data available mainly caused this.

C. RECOMMENDATIONS

My recommendation is that DLIFLC review and validate the performance objective tests to ensure that the tests measure the intended proficiency skills. With the models developed within this thesis, DLIFLC can predict success on test scores but each language utilizes different performance objectives with different degrees of error for each model.

APPENDIX A. EXAMPLE SPSS OUTPUT

LANG = Arabic

Variables Entered/Removed[a,b]

Model	Variables Entered	Variables Removed	Method
1	F2A	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	F7A	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	F1A	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
4	F5A	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
5	F3A	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
6	F10A	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
7	F9A	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: DLPT_L

b. LANG = Arabic

Model Summary[h,i]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.640[a]	.410	.409	4.17	.410	487.752	1	703	.000
2	.725[b]	.526	.524	3.74	.116	171.802	1	702	.000
3	.753[c]	.568	.566	3.57	.042	67.796	1	701	.000
4	.763[d]	.583	.580	3.51	.015	25.750	1	700	.000
5	.772[e]	.595	.592	3.46	.012	21.431	1	699	.000
6	.776[f]	.602	.599	3.43	.007	12.075	1	698	.001
7	.778[g]	.605	.601	3.42	.003	5.076	1	697	.025

a. Predictors: (Constant), F2A

b. Predictors: (Constant), F2A, F7A

c. Predictors: (Constant), F2A, F7A, F1A

d. Predictors: (Constant), F2A, F7A, F1A, F5A

e. Predictors: (Constant), F2A, F7A, F1A, F5A, F3A

f. Predictors: (Constant), F2A, F7A, F1A, F5A, F3A, F10A

g. Predictors: (Constant), F2A, F7A, F1A, F5A, F3A, F10A, F9A

h. Dependent Variable: DLPT_L

i. LANG = Arabic

Coefficients[a,b]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Correlations Partial	Correlations Part
1	(Constant)	29.643	.677		43.758	.000			
	F2A	.241	.011	.640	22.085	.000	.640	.640	.640
2	(Constant)	27.569	.628		43.907	.000			
	F2A	.169	.011	.448	15.029	.000	.640	.493	.391
	F7A	.106	.008	.391	13.107	.000	.611	.443	.341
3	(Constant)	27.708	.600		46.162	.000			
	F2A	.116	.013	.308	9.265	.000	.640	.330	.230
	F7A	8.754E-02	.008	.322	10.848	.000	.611	.379	.269
	F1A	9.447E-02	.011	.275	8.234	.000	.631	.297	.205
4	(Constant)	27.228	.597		45.575	.000			
	F2A	.105	.013	.278	8.386	.000	.640	.302	.205
	F7A	6.407E-02	.009	.236	6.979	.000	.611	.255	.170
	F1A	8.383E-02	.011	.244	7.309	.000	.631	.266	.178
	F5A	5.262E-02	.010	.176	5.074	.000	.606	.188	.124
5	(Constant)	25.720	.673		38.217	.000			
	F2A	.108	.012	.287	8.756	.000	.640	.314	.211
	F7A	4.861E-02	.010	.179	5.039	.000	.611	.187	.121
	F1A	8.253E-02	.011	.240	7.298	.000	.631	.266	.176
	F5A	4.788E-02	.010	.160	4.660	.000	.606	.174	.112
	F3A	3.401E-02	.007	.130	4.629	.000	.414	.172	.111
6	(Constant)	23.859	.856		27.872	.000			
	F2A	.101	.012	.268	8.142	.000	.640	.295	.194
	F7A	4.648E-02	.010	.171	4.846	.000	.611	.180	.116
	F1A	7.704E-02	.011	.224	6.799	.000	.631	.249	.162
	F5A	4.593E-02	.010	.154	4.499	.000	.606	.168	.107
	F3A	2.856E-02	.007	.109	3.830	.000	.414	.143	.091
	F10A	4.021E-02	.012	.098	3.475	.001	.472	.130	.083
7	(Constant)	27.079	1.665		16.266	.000			
	F2A	.104	.012	.275	8.333	.000	.640	.301	.198
	F7A	5.006E-02	.010	.184	5.164	.000	.611	.192	.123
	F1A	7.908E-02	.011	.230	6.977	.000	.631	.255	.166
	F5A	4.455E-02	.010	.149	4.369	.000	.606	.163	.104
	F3A	2.740E-02	.007	.104	3.677	.000	.414	.138	.088
	F10A	3.847E-02	.012	.094	3.327	.001	.472	.125	.079
	F9A	-3.69E-02	.016	-.056	-2.253	.025	.116	-.085	-.054

a. Dependent Variable: DLPT_L

b. LANG = Arabic

Residuals Statistics[a,b]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	32.00	55.56	44.19	4.20	712
Residual	-13.68	10.78	-8.86E-03	3.40	712
Std. Predicted Value	-2.892	2.695	-.001	.997	712
Std. Residual	-3.995	3.150	-.003	.993	712

a. Dependent Variable: DLPT_L

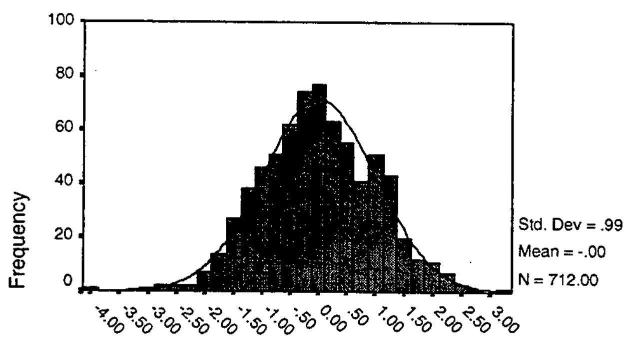
b. LANG = Arabic

Charts

Histogram

Dependent Variable: DLPT_L

LANG: AD Arabic

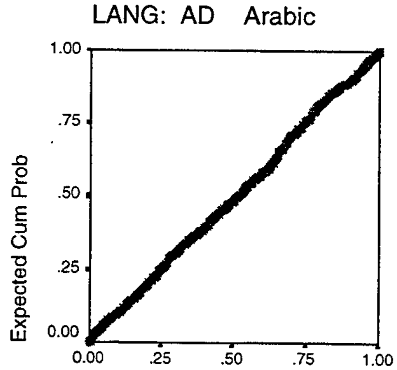


Regression Standardized Residual

Normal P-P Plot of Regression Sta

Dependent Variable: DLPT_L

LANG: AD Arabic

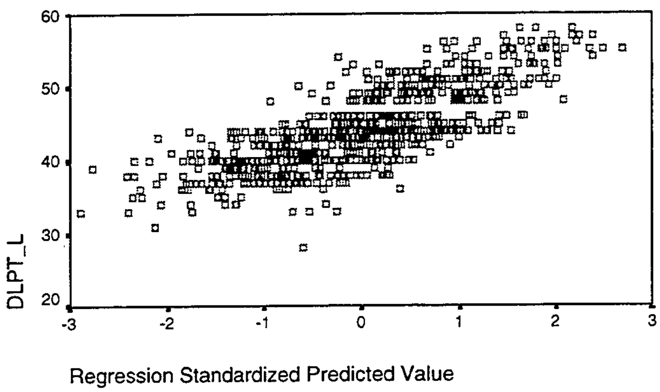


Observed Cum Prob

Scatterplot

Dependent Variable: DLPT_L

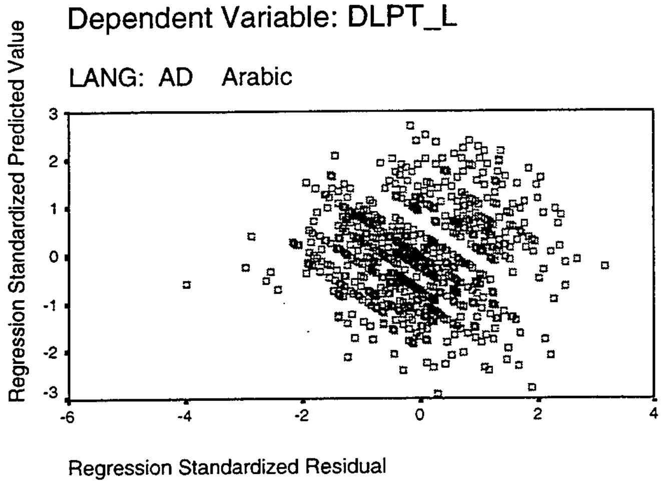
LANG: AD Arabic



Scatterplot

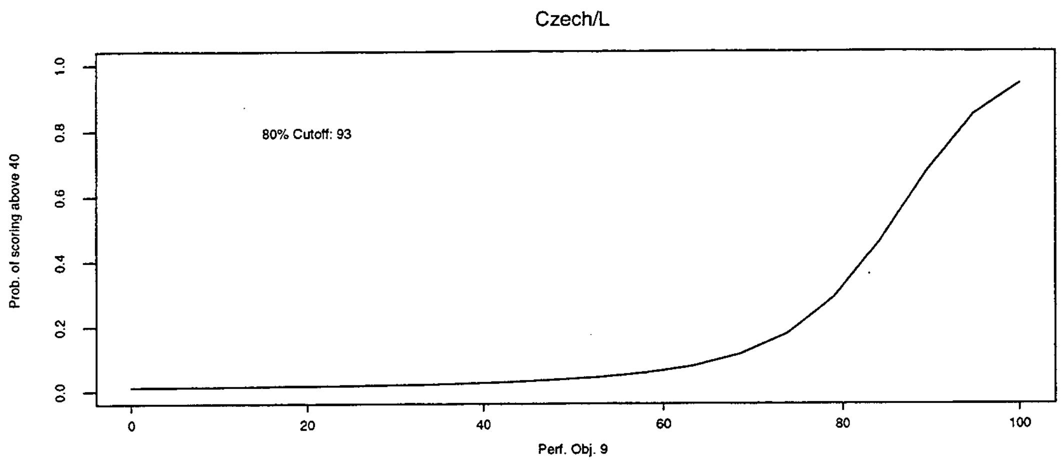
Dependent Variable: DLPT_L

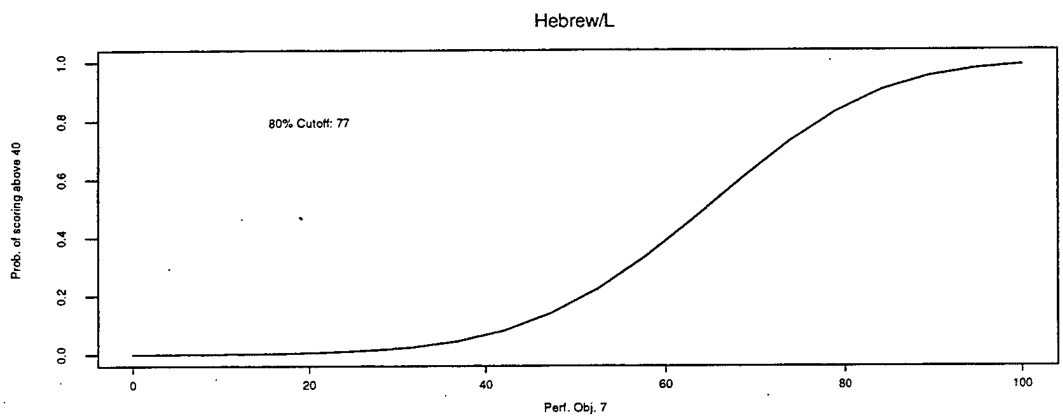
LANG: AD Arabic

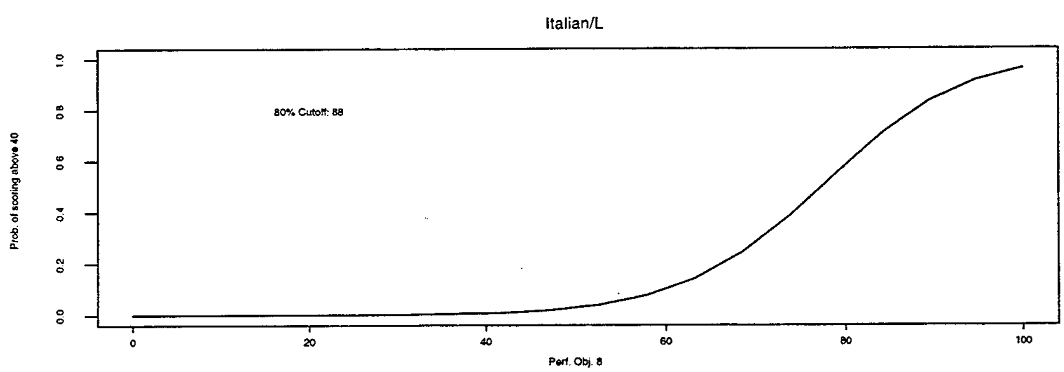


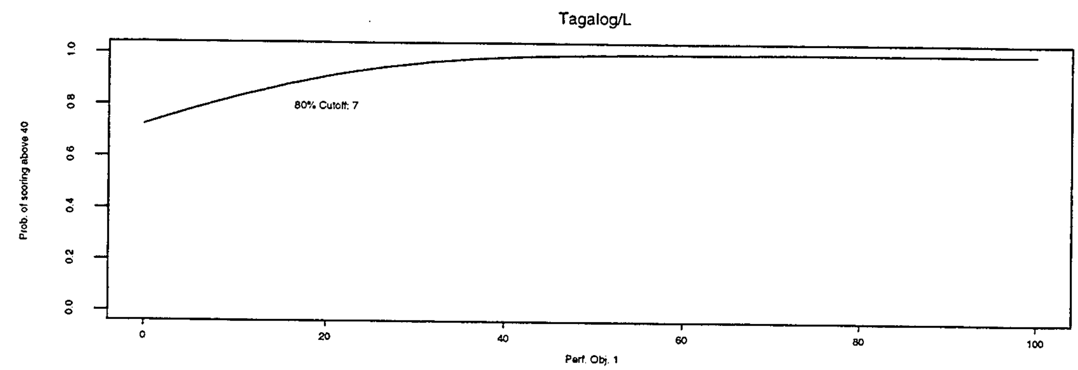
APPENDIX B. PROBABILITY CHARTS FOR DLPT FOR ALL MODELS WITH ONE MAIN EFFECT

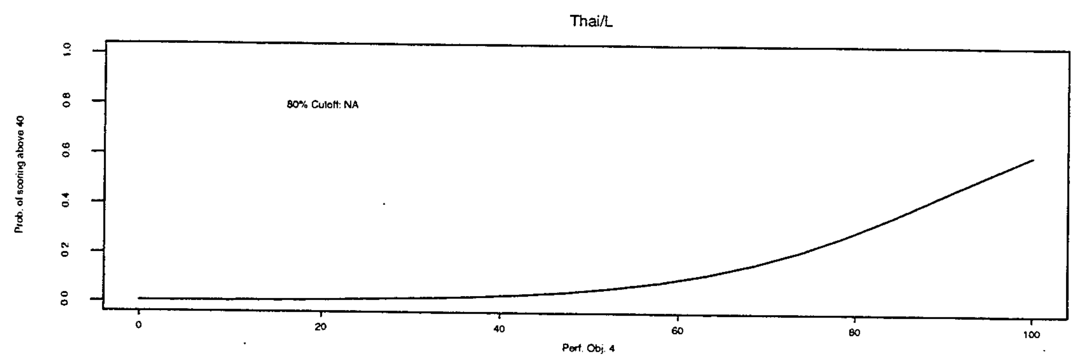
A. DLPT_L



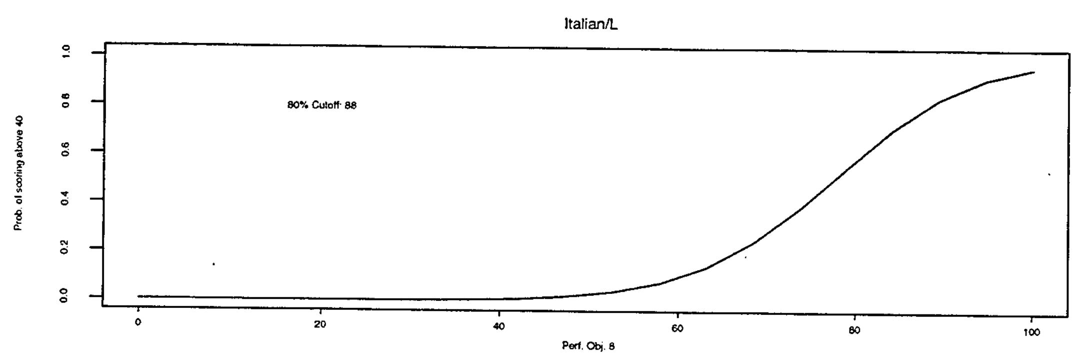


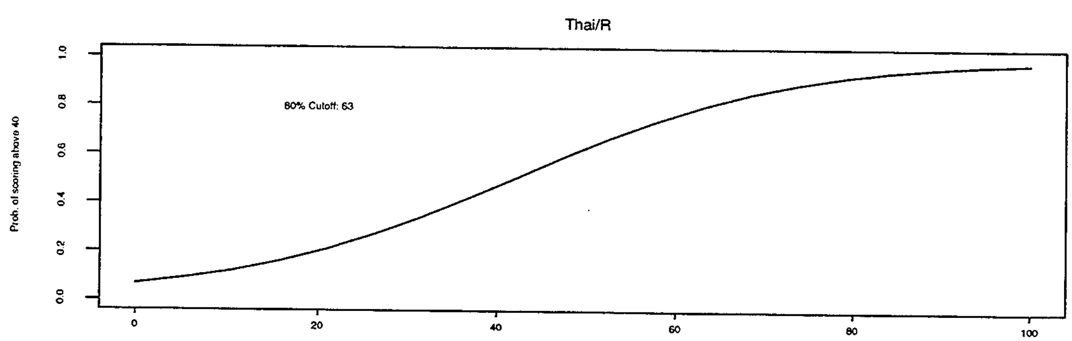






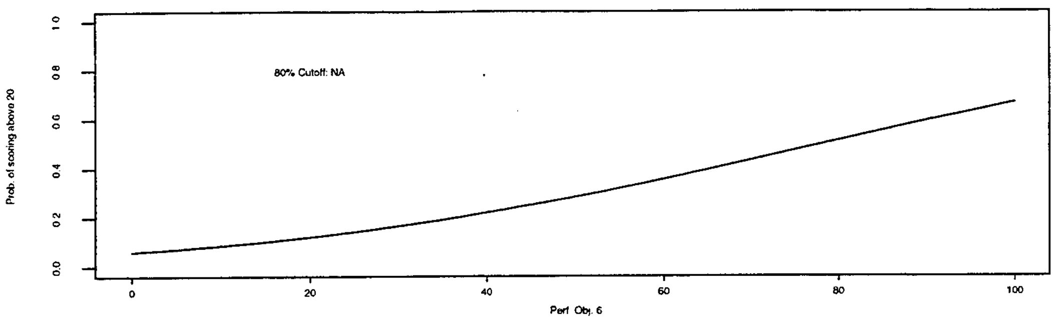
B. DLPT_R



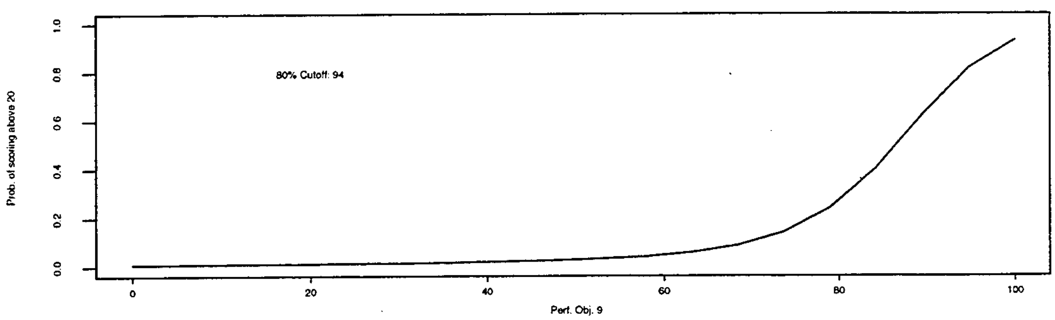


C. DLPT_S

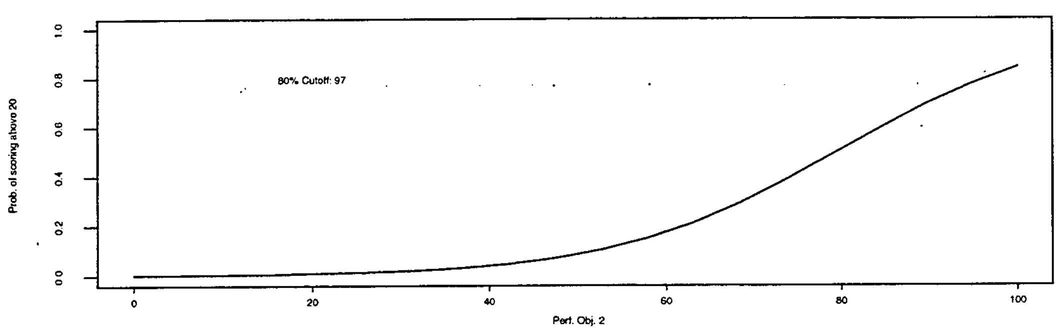
Chinese-Mandarin/S

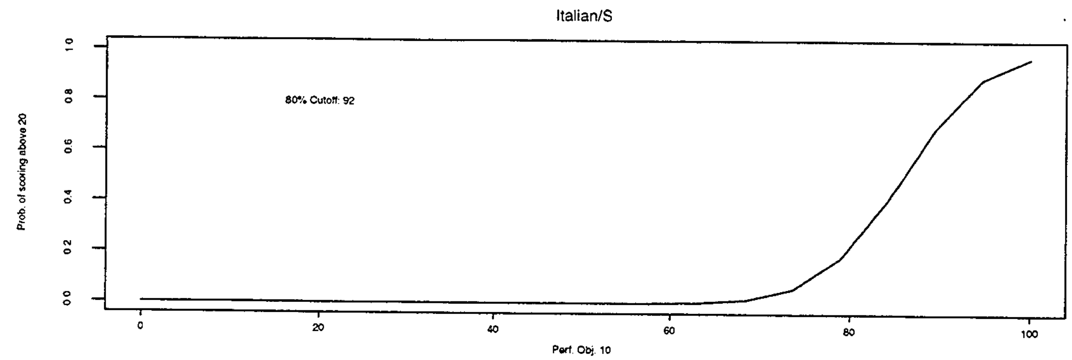


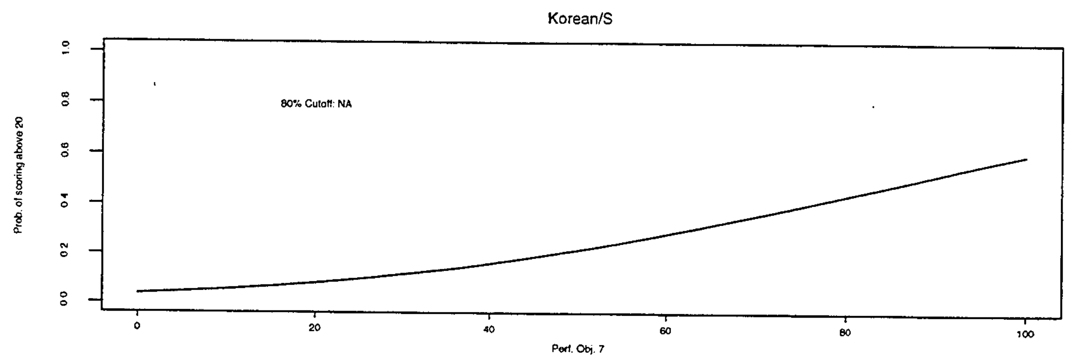
Czech/S

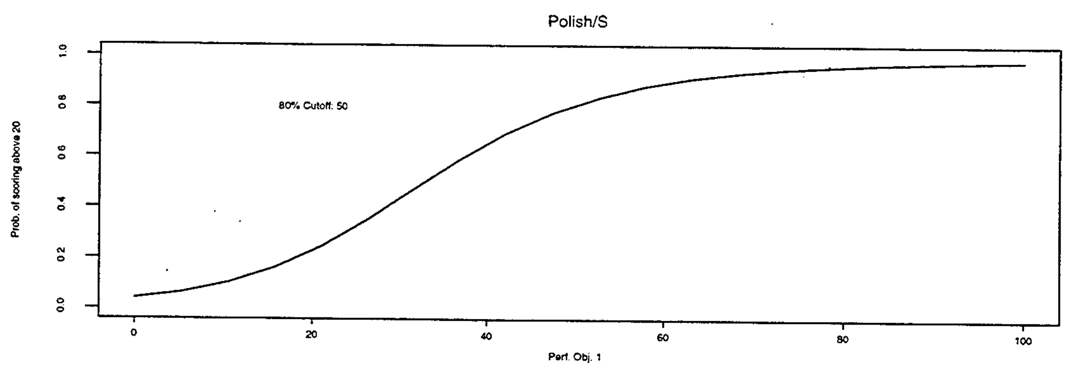


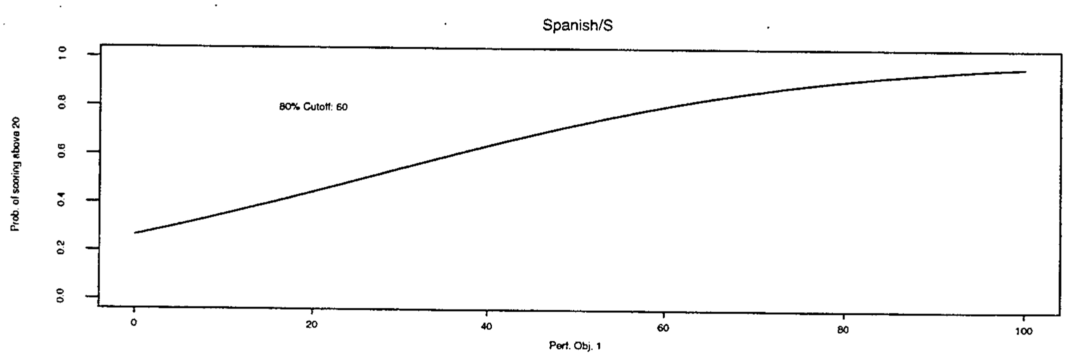
Hebrew/S

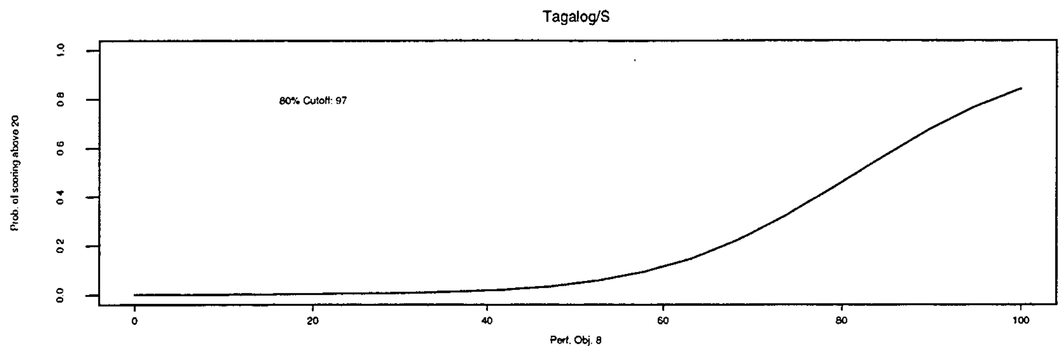



Italian/S
80% Cutoff: 92
Prob. of scoring above 20
Perf. Obj. 10


Korean/S
80% Cutoff: NA
Prob. of scoring above 20
Perf. Obj. 7


Polish/S
80% Cutoff: 50
Prob. of scoring above 20
Perf. Obj. 1

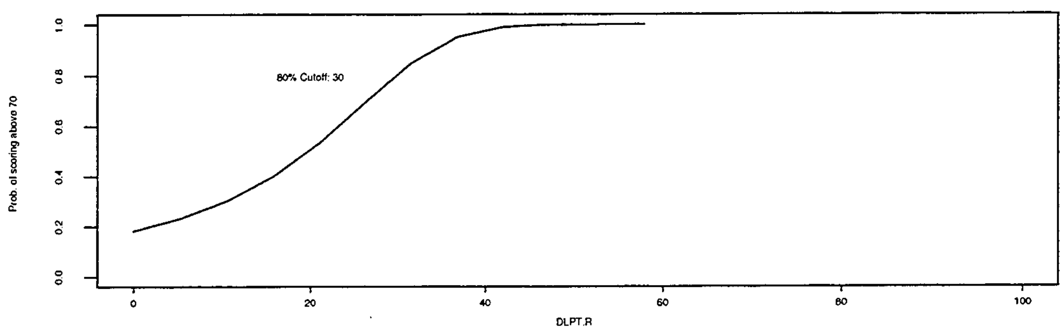

Spanish/S
80% Cutoff: 60
Prob. of scoring above 20
Perf. Obj. 1


Tagalog/S
80% Cutoff: 97
Prob. of scoring above 20
Perf. Obj. 8
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100

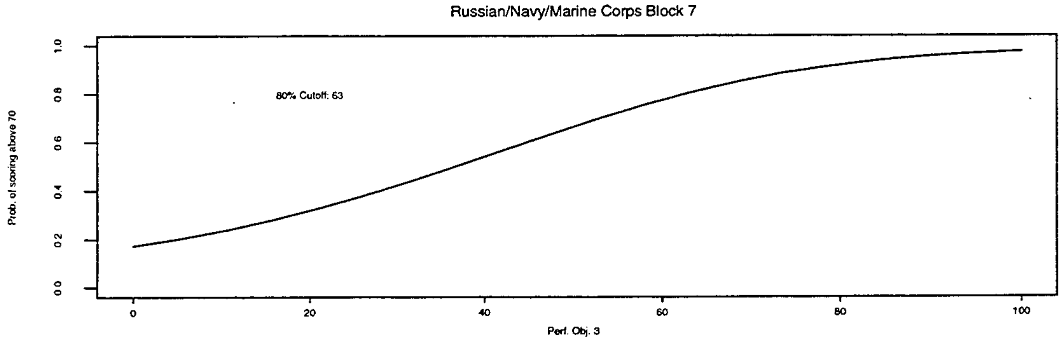
APPENDIX C. PROBABILITY CHARTS FOR "BLOCK" TESTS OF RUSSIAN GAFB MODELS WITH SINGLE MAIN EFFECTS

A. NAVY/MARINE CORPS

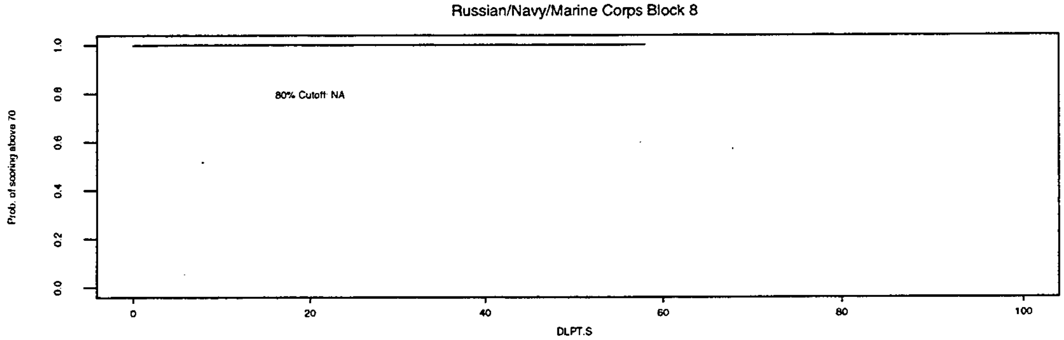
Russian/Navy/Marine Corps Block 3



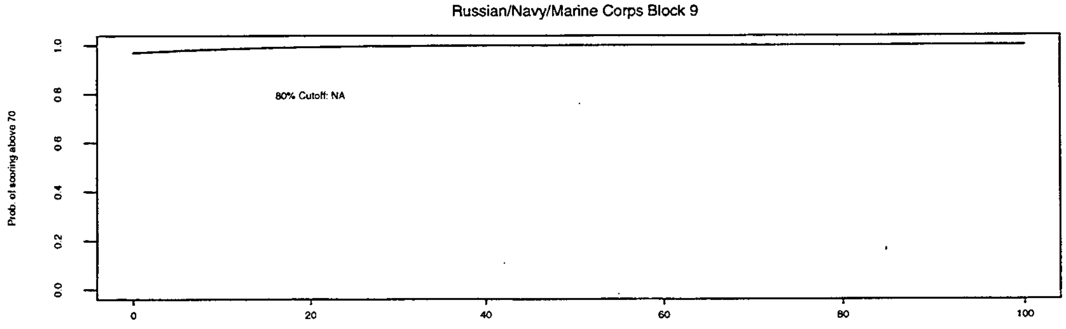
Russian/Navy/Marine Corps Block 7

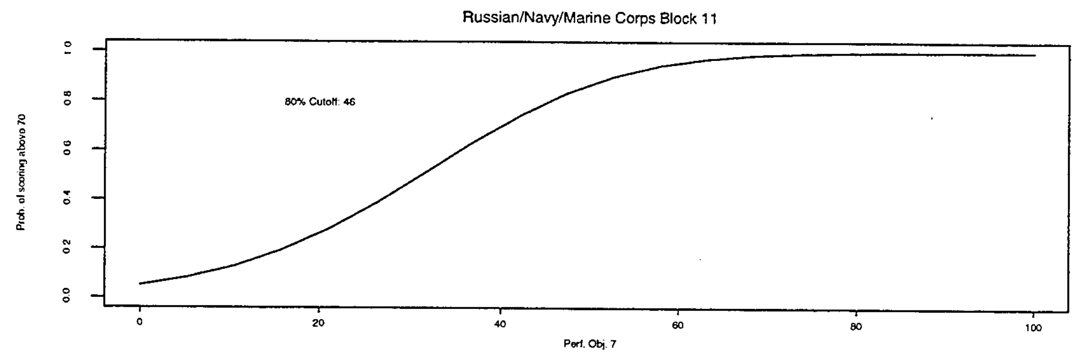


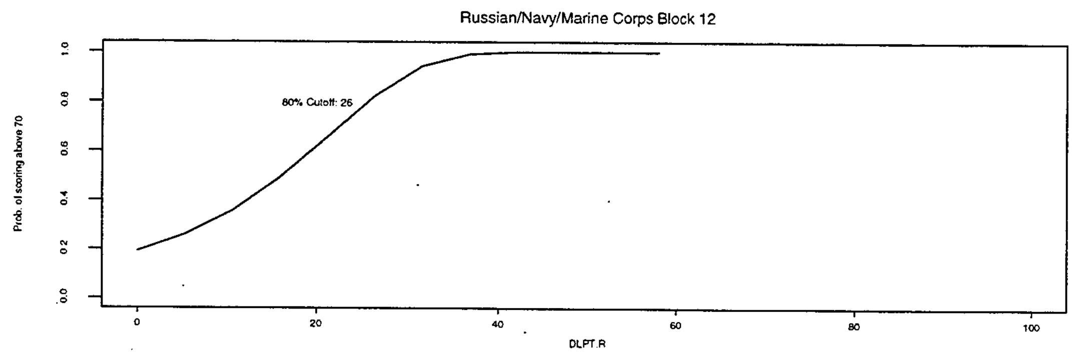
Russian/Navy/Marine Corps Block 8

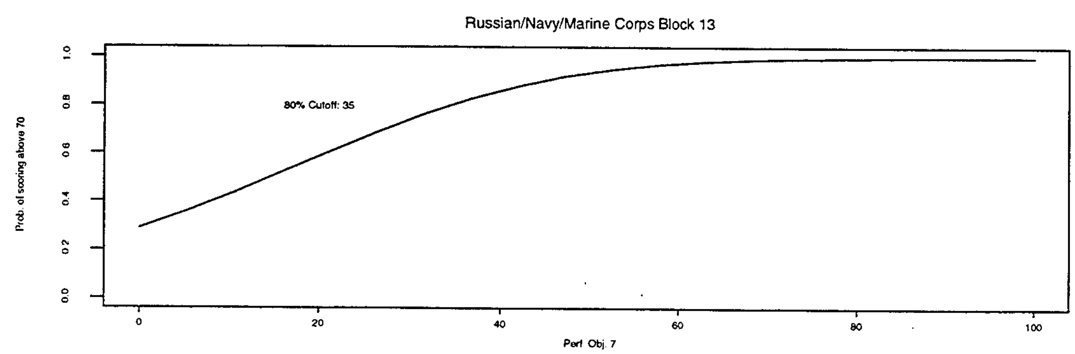


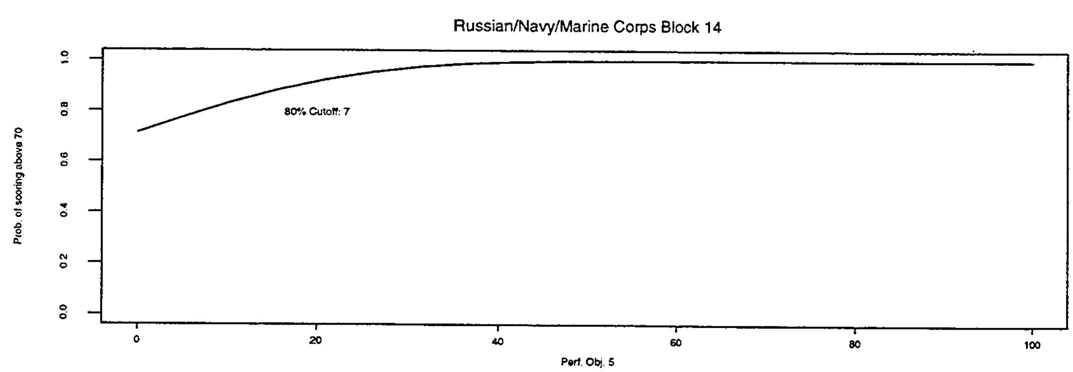
Russian/Navy/Marine Corps Block 9

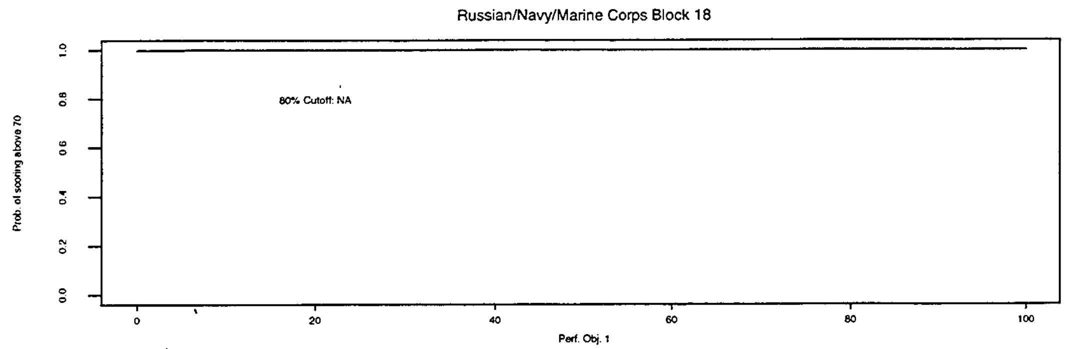


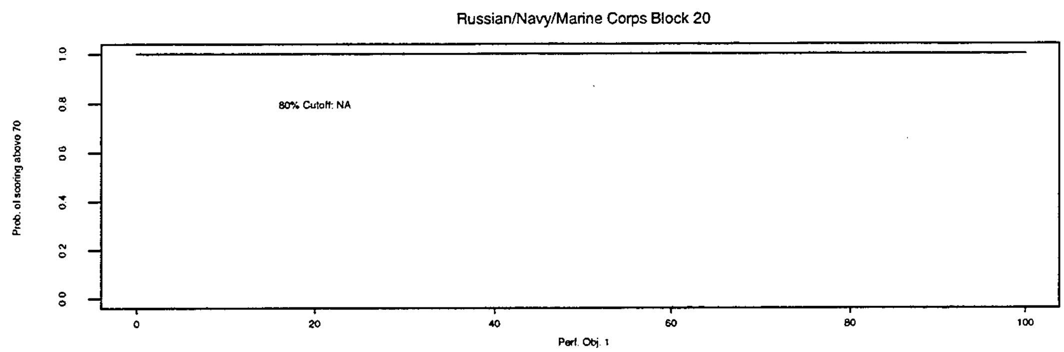

Russian/Navy/Marine Corps Block 11
80% Cutoff: 46
Prob. of scoring above 70
Perf. Obj. 7

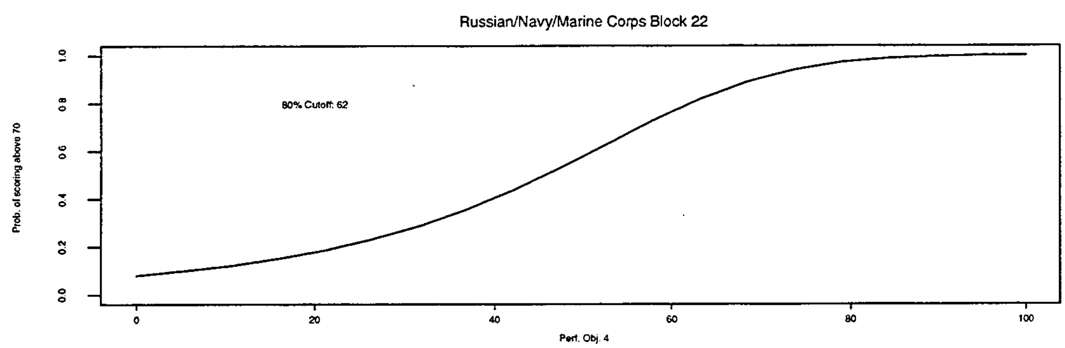

Russian/Navy/Marine Corps Block 12
80% Cutoff: 26
Prob. of scoring above 70
DLPT.R


Russian/Navy/Marine Corps Block 13
80% Cutoff: 35
Prob. of scoring above 70
Perf Obj. 7

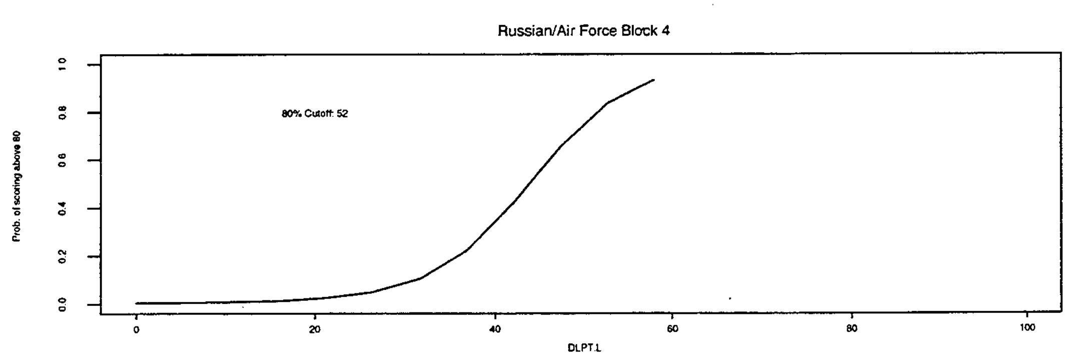

Russian/Navy/Marine Corps Block 14
80% Cutoff: 7
Prob. of scoring above 70
Perf. Obj. 5

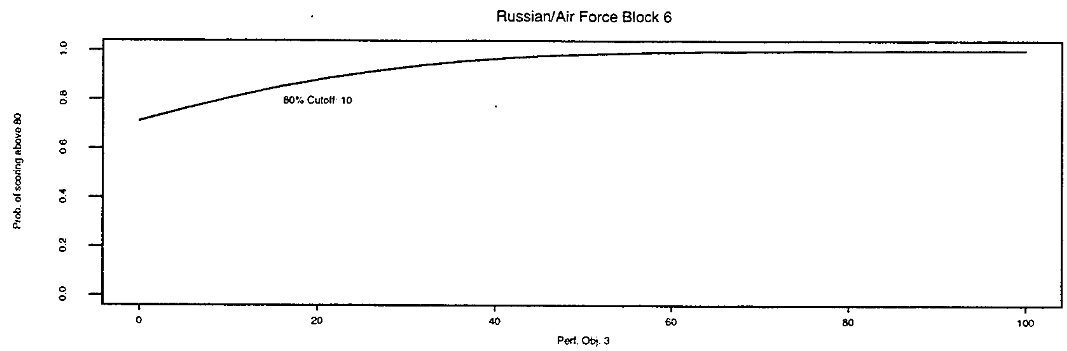

Russian/Navy/Marine Corps Block 18
80% Cutoff: NA
Prob. of scoring above 70
Perf. Obj. 1

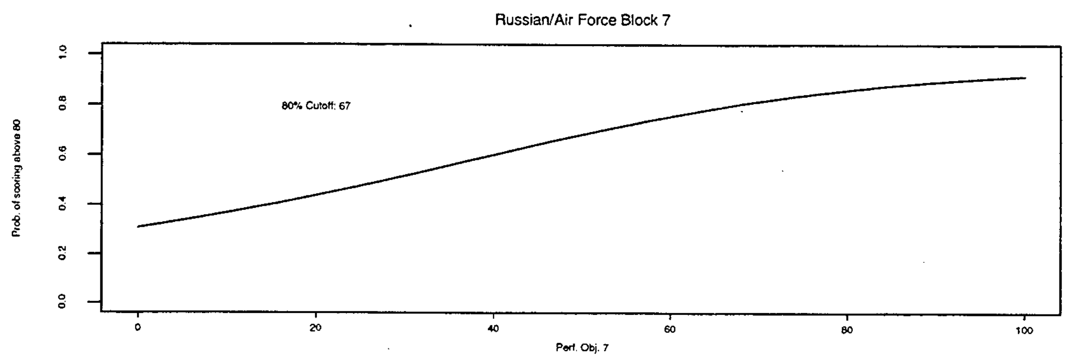

Russian/Navy/Marine Corps Block 20
80% Cutoff: NA
Prob. of scoring above 70
Perf. Obj. 1

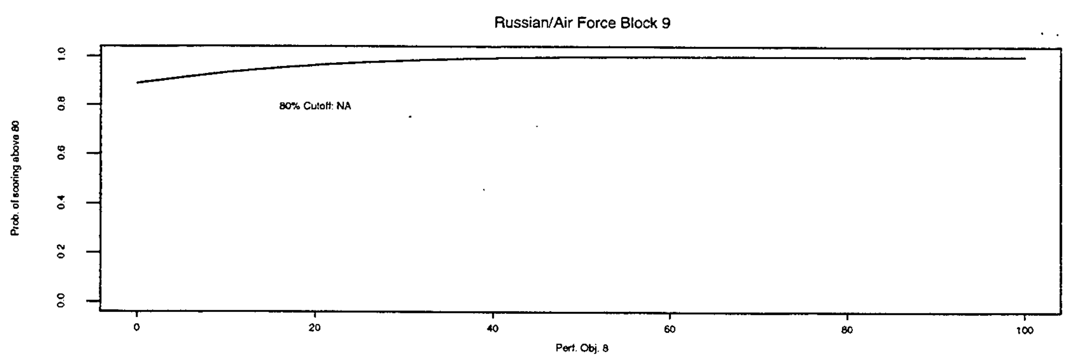

Russian/Navy/Marine Corps Block 22
80% Cutoff: 62
Prob. of scoring above 70
Perf. Obj. 4

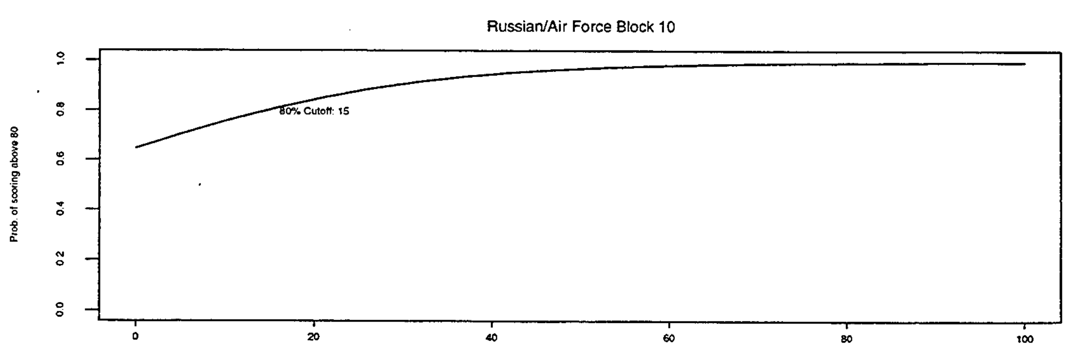
B. AIR FORCE


Russian/Air Force Block 4
80% Cutoff: 52
Prob. of scoring above 80
DLPT L

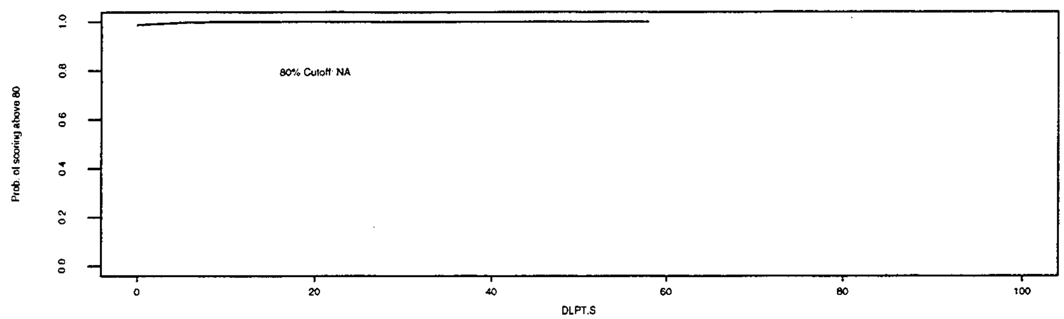

Russian/Air Force Block 6
Prob. of scoring above 80
80% Cutoff: 10
Perf. Obj. 3


Russian/Air Force Block 7
Prob. of scoring above 80
80% Cutoff: 67
Perf. Obj. 7

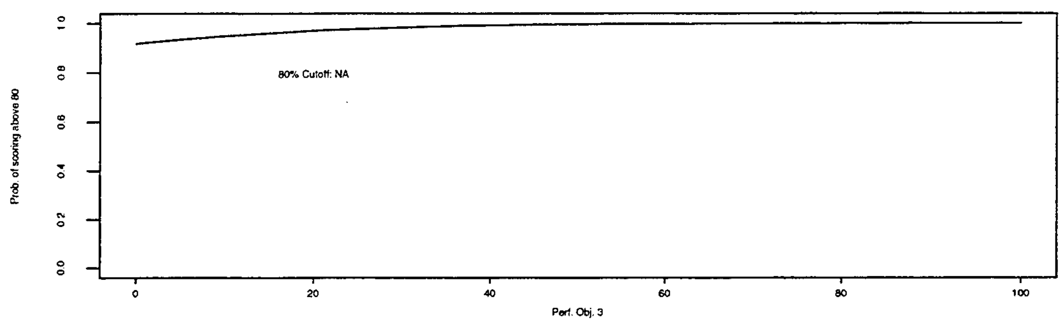

Russian/Air Force Block 9
Prob. of scoring above 80
80% Cutoff: NA
Perf. Obj. 8


Russian/Air Force Block 10
Prob. of scoring above 80
80% Cutoff: 15
Perf. Obj. 1

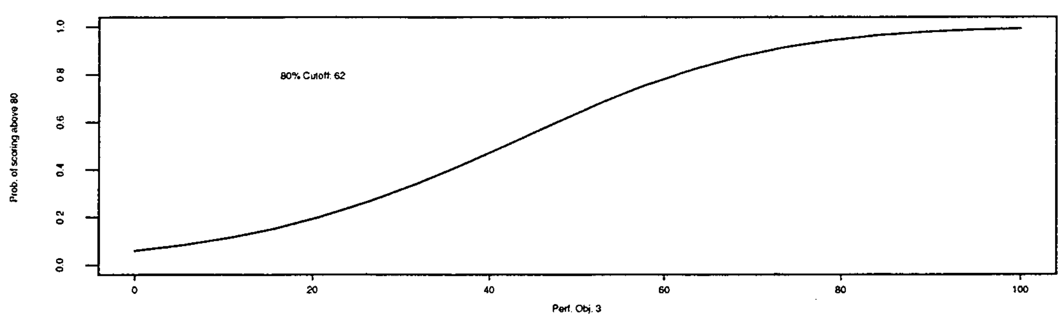
Russian/Air Force Block 11



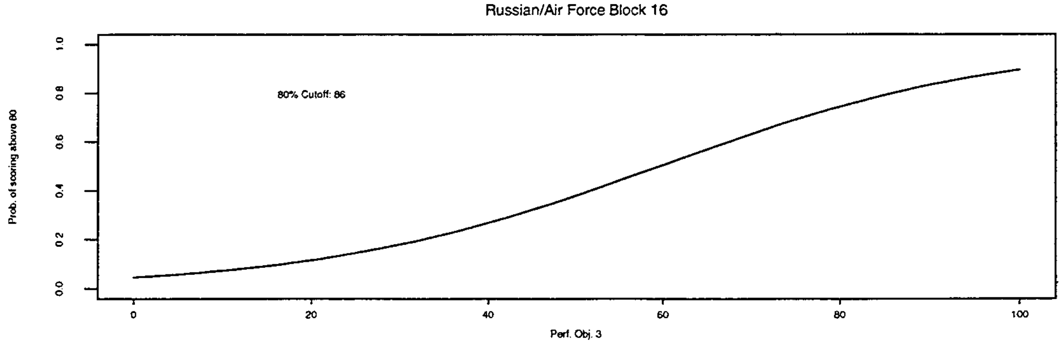
Russian/Air Force Block 12

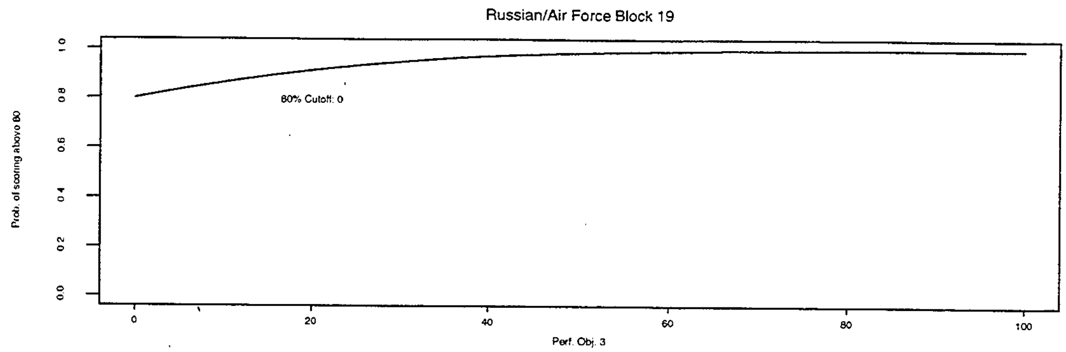


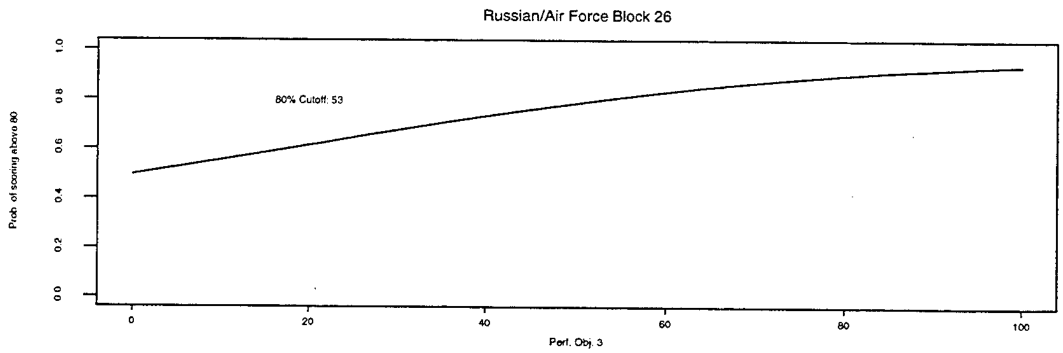
Russian/Air Force Block 14



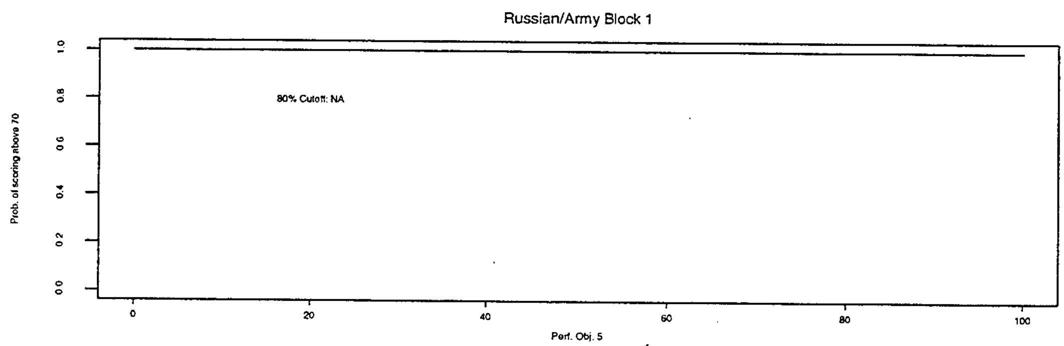
Russian/Air Force Block 16

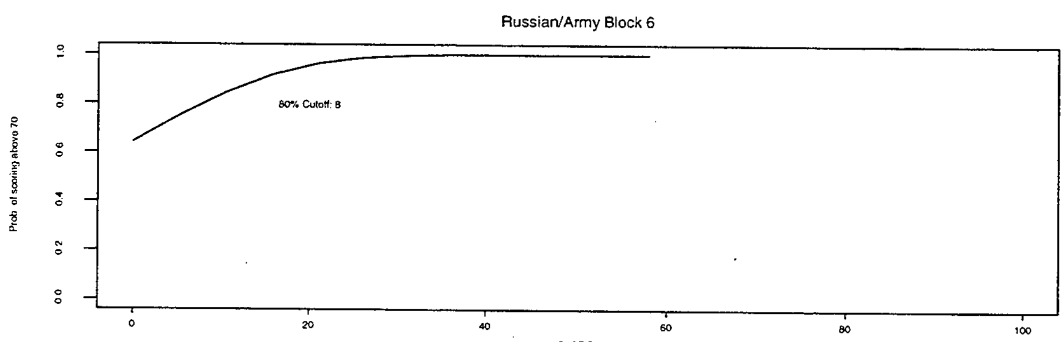


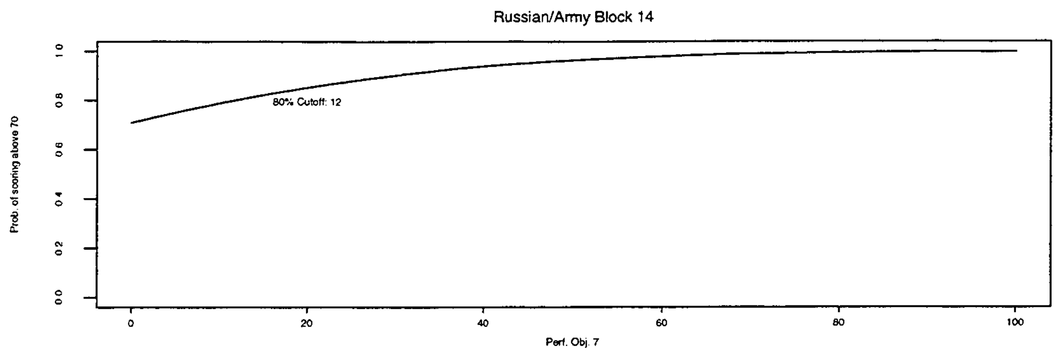

Russian/Air Force Block 19
Prob. of scoring above 80
80% Cutoff: 0
Perf. Obj. 3
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100


Russian/Air Force Block 26
Prob. of scoring above 80
80% Cutoff: 53
Perf. Obj. 3
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100

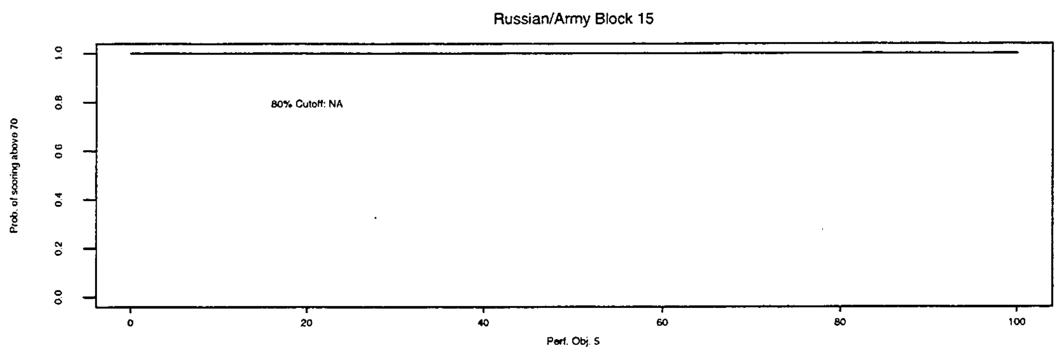
C. ARMY


Russian/Army Block 1
Prob. of scoring above 70
80% Cutoff: NA
Perf. Obj. 5
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100


Russian/Army Block 6
Prob. of scoring above 70
80% Cutoff: 8
DLPT.R
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100



Russian/Army Block 14
80% Cutoff: 12
Prob. of scoring above 70
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
Perf. Obj. 7



Russian/Army Block 15
80% Cutoff: NA
Prob. of scoring above 70
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
Perf. Obj. 5

APPENDIX D. S-PLUS FUNCTION FOR PROBABILITY GRAPHS

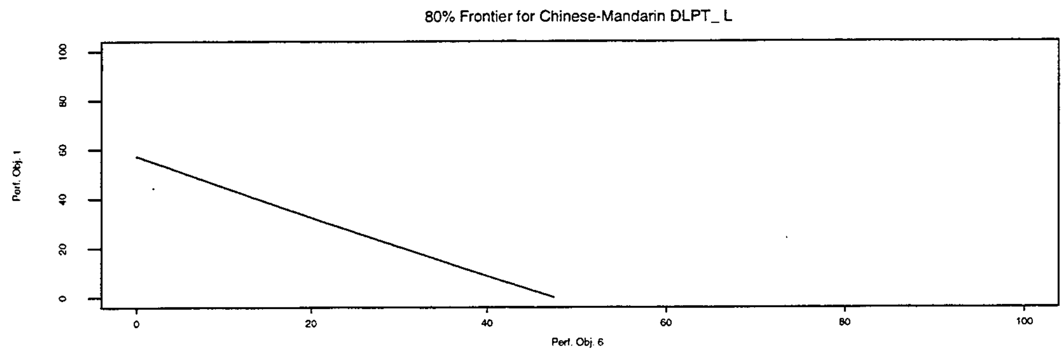
```
function(lang, first, dlpt, data = big, crit = 40, prob = 0.8,
return.model = F, n = 20)
{
#
# jt3: Do two-d "prob.  of passing | FLO") plot
#
# Arguments: lang: two-letter language abbreviation
#            first: name of first FLO test
#             dltp: one-letter choice of dlpt
# Start by trying to handle zeros
      zeros <- data[, first] == 0    #
#
# Stick these things into frame 1.  Don't ask.
#
      assign("lang", lang, frame = 1)
      assign("zeros", zeros, frame = 1)    #
#
# Create the text of the model statement, and execute it.
#
      model.txt <- paste("lm(DLPT.", dlpt, " ~ ", first,
            ", data = data, na.action = na.omit, subset = LANG == lang &
!zeros)", sep = "")
      out <- eval(parse(text = model.txt))       #
#
# Set up a vector of FLOs in result[,1].  For each element in the
vector, find the predicted
# DLPT score and the associated SE of prediction.  Then compute the
probability of
# passing the test.
#
      result <- matrix(0, n, 2)
      result[, 1] <- seq(0, 100, length = n)
      preds <- predict(out, cbind(1, result[, 1]), se.fit = T)
      sds <- sqrt(preds$resid^2 + preds$se^2)
      result[, 2] <- 1 - pnorm((crit - preds$fit)/sds)       #
#
# Extract FLO number (10 is a special case) for the label.
#
      if(nchar(first) == 4)
            fx <- substring(first, 2, 3)
      else fx <- substring(first, 2, 2)    #
      plot(result[, 1], result[, 2], ylim = c(0, 1), type = "l", xlab =
paste("Perf.  Obj.", fx), ylab =
            paste("Prob.  of scoring above", crit), main =
paste(xref[xref[, "two"] == lang, "long"], "/",
            dlpt, sep = ""))  #
#
# ...then compute and display the cut-off itself.
#
      app <- approx(result[, 2], result[, 1], 0.8)
      text(20, 0.8, paste("80% Cutoff:", round(app$y)))
      return(result)
}
```

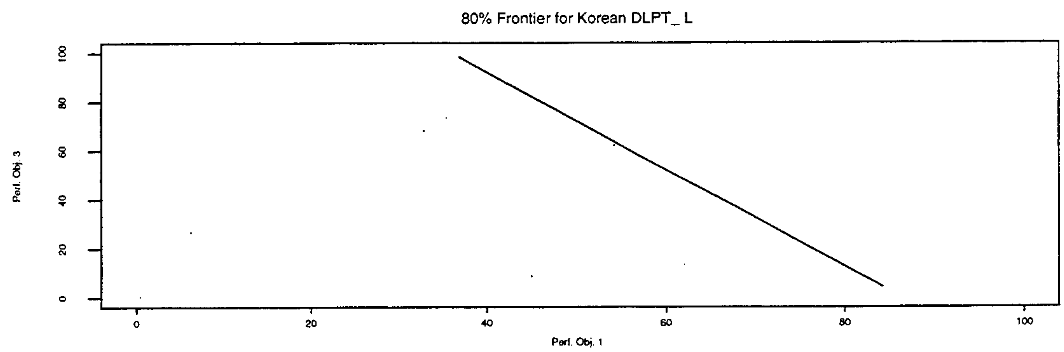
APPENDIX E. S-PLUS FUNCTION FOR FRONTIER GRAPHS

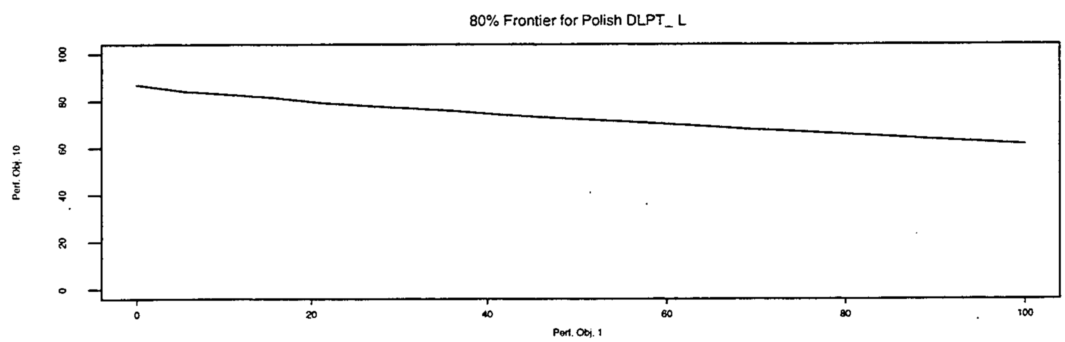
```
function(lang, first, second, dlpt, data = big, crit = 40, prob = 0.8,
return.model = F, n = 20)
{
#
# Prepare.grid: prepare a grid for making a cool 3D plot.
#
# Arguments: lang: two-letter language abbreviation
#           first: name of first FLO test
#          second: name of second FLO test
#            dlpt: one-letter choice of dlpt (L, R, or S)
#            crit: cut-off value of interest
#            prob: probability of exceeding "crit" on "dlpt"
#    return.model: If True, return model: useful for debugging
#               n: Number of points at which to compute prob.
#
# Start by trying to handle zeros
      zeros <- data[, first] == 0 | data[, second] == 0     #
#
# Stick these things into frame 1.  This gets around a well-known bug
# in Splus in which modelling functions cannot find objects in local
frames.
#
      assign("lang", lang, frame = 1)
      assign("zeros", zeros, frame = 1)    #
#
# Create the text of the model statement, and execute it.  Save it in
"out."
#
      model.txt <- paste("lm(DLPT.", dlpt, " ~ ", first, " + ", second,
            ", data = data, na.action = na.omit, subset = LANG == lang &
!zeros)", sep = "")
      out <- eval(parse(text = model.txt))      #
#
# Set up the matrix of results.  The first column is the x's.
#
      result <- matrix(0, n, 2)
      result[, 1] <- seq(0, 100, length = n)     #
#
# Set up the x-label.  "F10A" is a special case.  These might be DLPT's,
too,
# for the GAFB case.
#
      if(substring(first, 1, 1) == "F") {
            if(nchar(first) == 4)
                  f.txt <- paste("Perf.  Obj.", substring(first, 2, 3))
            else f.txt <- paste("Perf.  Obj.", substring(first, 2, 2))
      }
      else f.txt <- first
      if(substring(second, 1, 1) == "F") {
            if(nchar(second) == 4)
                  s.txt <- paste("Perf.  Obj.", substring(second, 2, 3))
            else s.txt <- paste("Perf.  Obj.", substring(second, 2, 2))
      }
      else s.txt <- second
      for(i in 1:n) {
            cat("Finding frontier ", i, "\n")
            second.test <- seq(0, 100, length = n)
```

```
		pred.list <- predict(out, cbind(1, rep(result[i, 1], n),
second.test), se.fit = T)
		preds <- pred.list$fit
		sds <- sqrt(pred.list$resid^2 + pred.list$se^2)
		temp.res <- 1 - pnorm((crit - preds)/sds)
		app.out <- approx(temp.res, second.test, 0.8)
		result[i, 2] <- app.out$y
	}
#
#
# Draw the picture and quit.
#
	plot(result[, 1], result[, 2], xlab = f.txt, ylab = s.txt, main =
paste("80% Frontier for", xref[xref[,
		"two"] == lang, "long"], "DLPT_", dlpt), type = "l", xlim =
c(0, 100), ylim = c(0, 100))
	if(return.model == T)
		return(out)
}
```

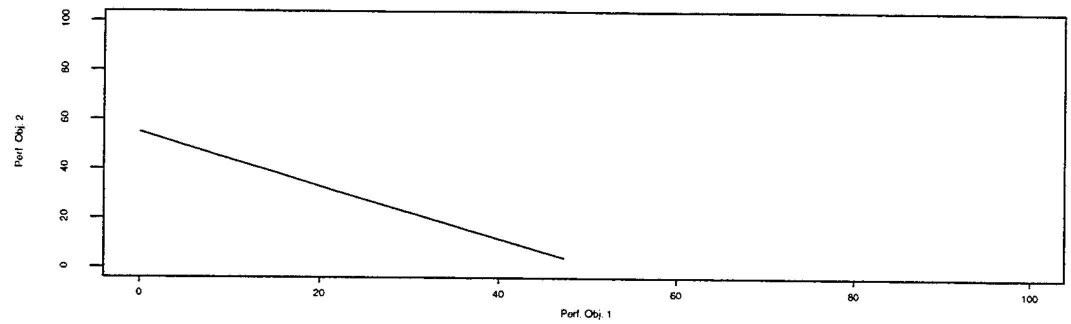
APPENDIX F. FRONTIER GRAPHS FOR DLIFLC MODELS WITH TWO MAIN EFFECTS



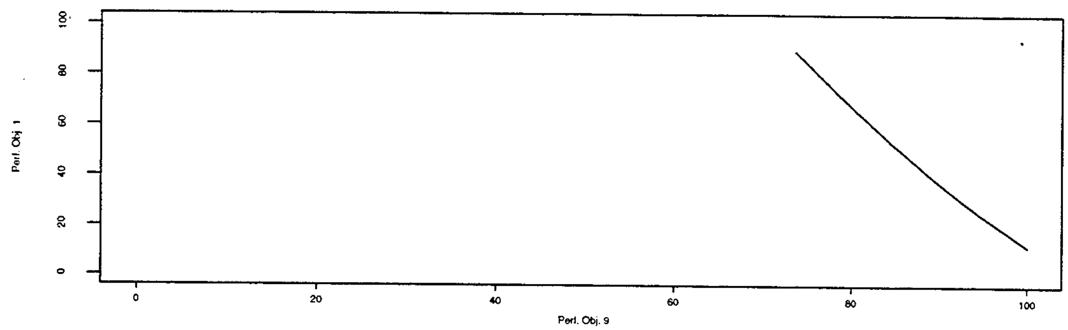




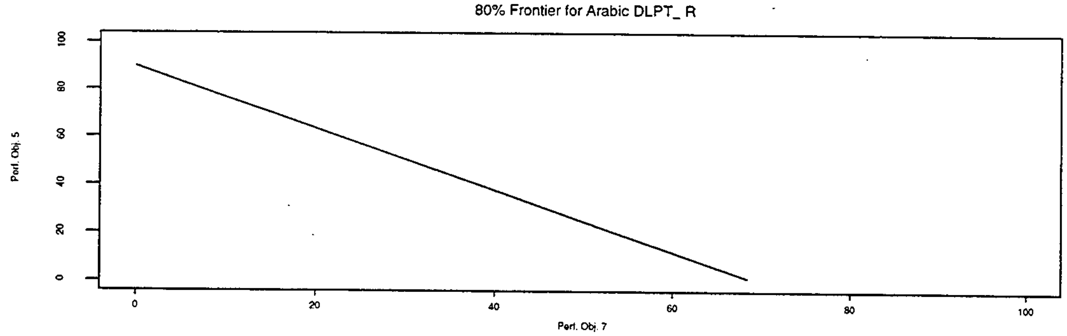
80% Frontier for Russian DLPT_ L



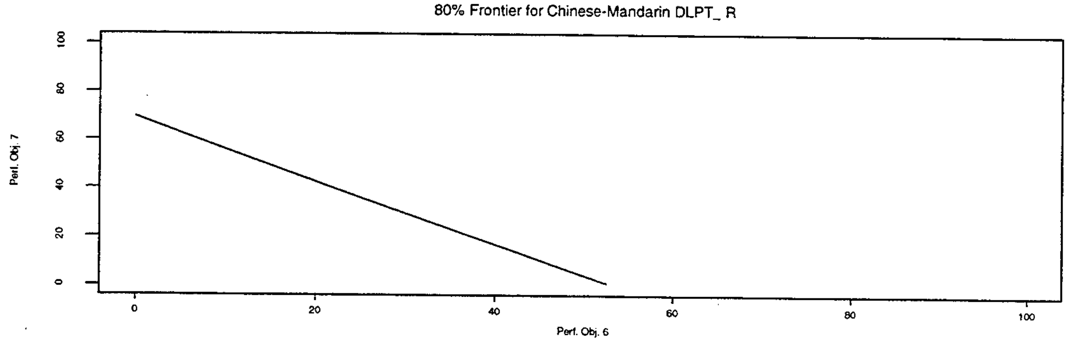
80% Frontier for Vietnamese DLPT_ L

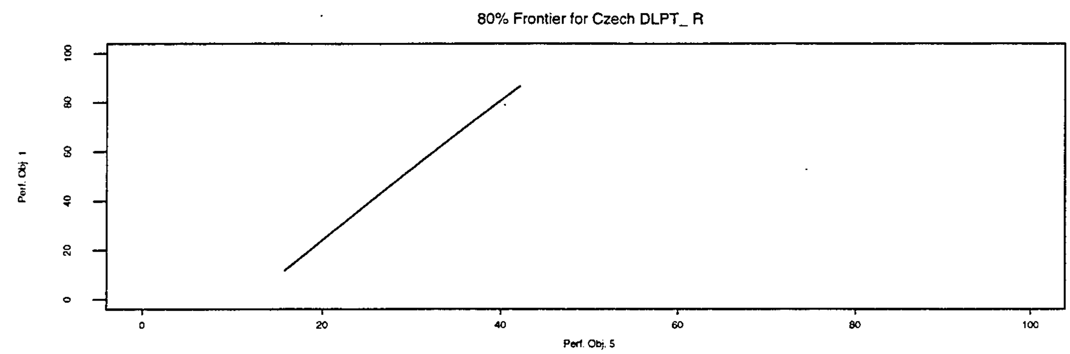


80% Frontier for Arabic DLPT_ R

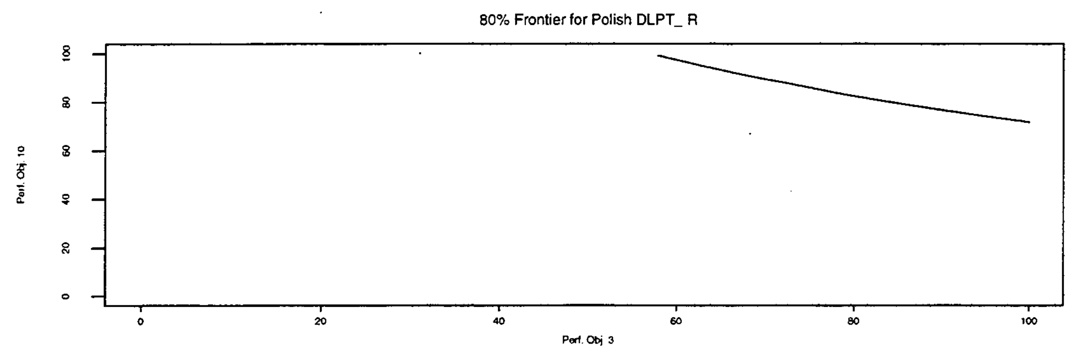


80% Frontier for Chinese-Mandarin DLPT_ R

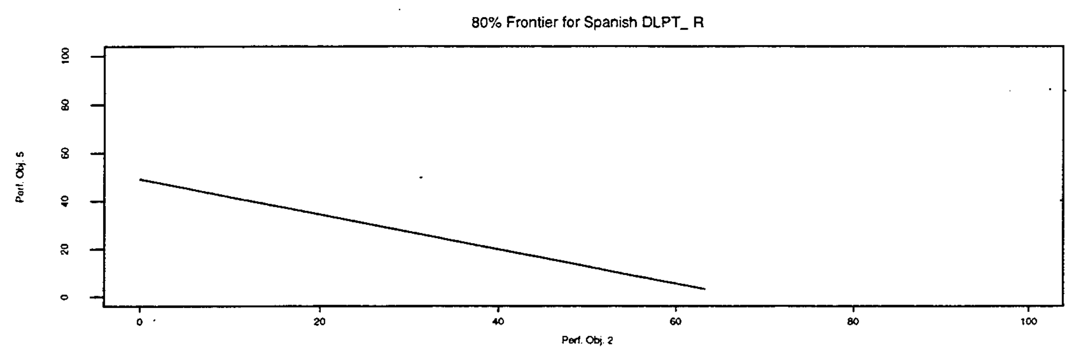




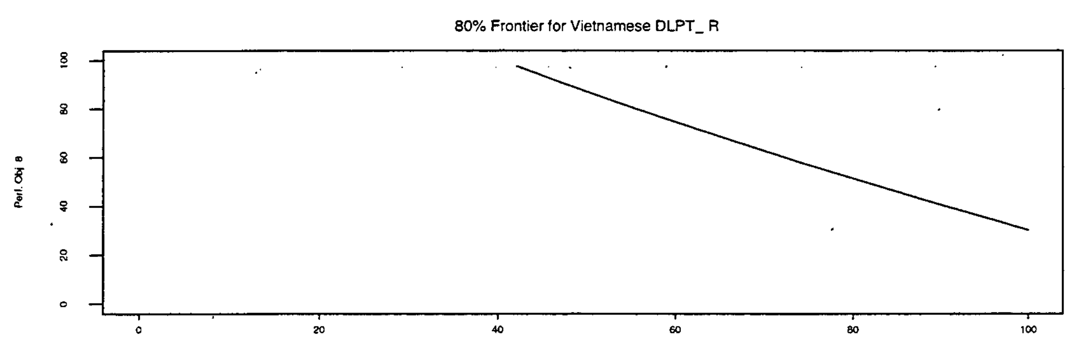
80% Frontier for Czech DLPT_ R
Perf. Obj 1
Perf. Obj. 5



80% Frontier for Polish DLPT_ R
Perf. Obj 10
Perf. Obj 3

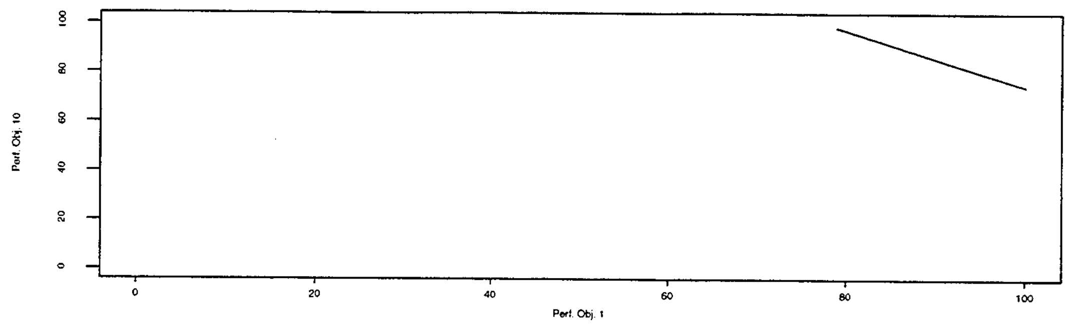


80% Frontier for Spanish DLPT_ R
Perf. Obj. 5
Perf. Obj. 2

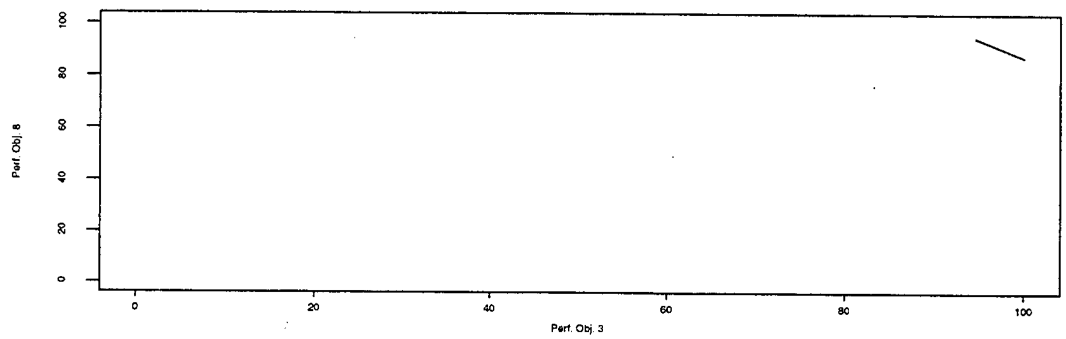


80% Frontier for Vietnamese DLPT_ R
Perf. Obj 8
Perf. Obj. 5

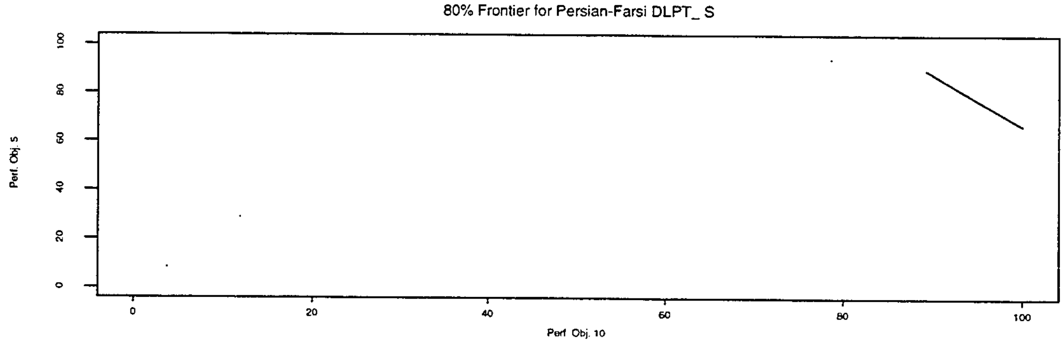
80% Frontier for Arabic DLPT_ S



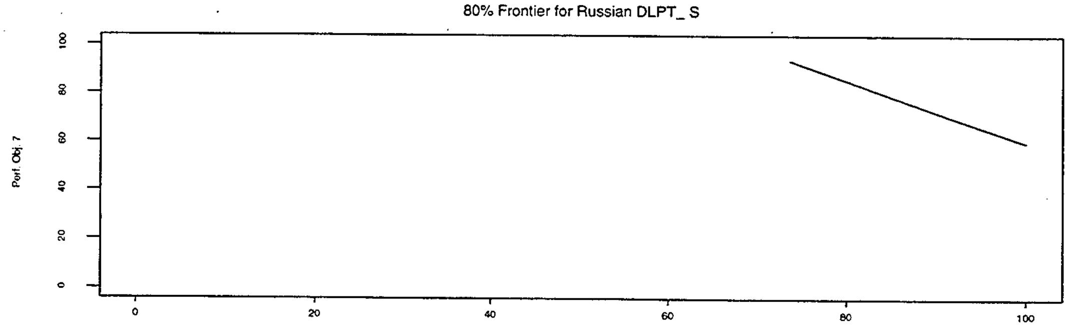
80% Frontier for French DLPT_ S



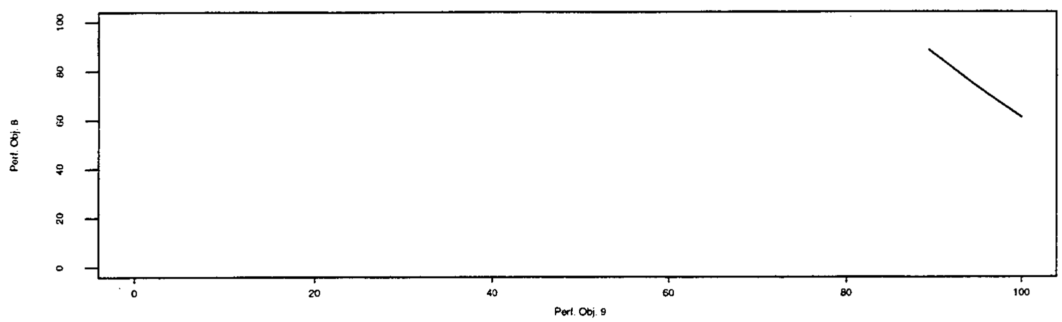
80% Frontier for Persian-Farsi DLPT_ S



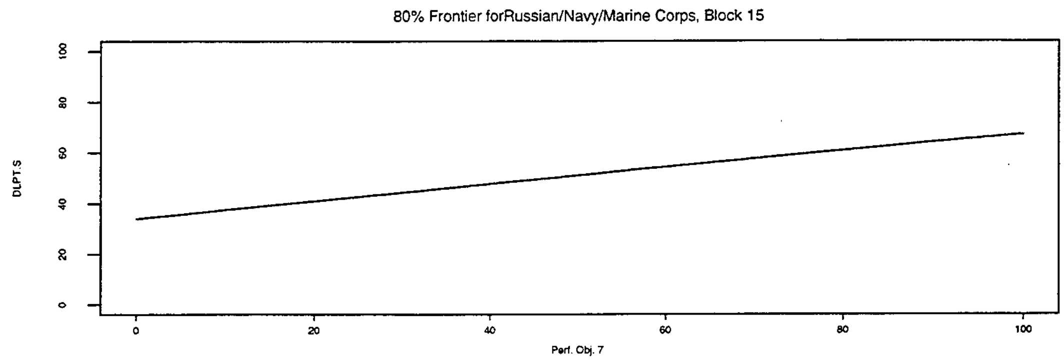
80% Frontier for Russian DLPT_ S

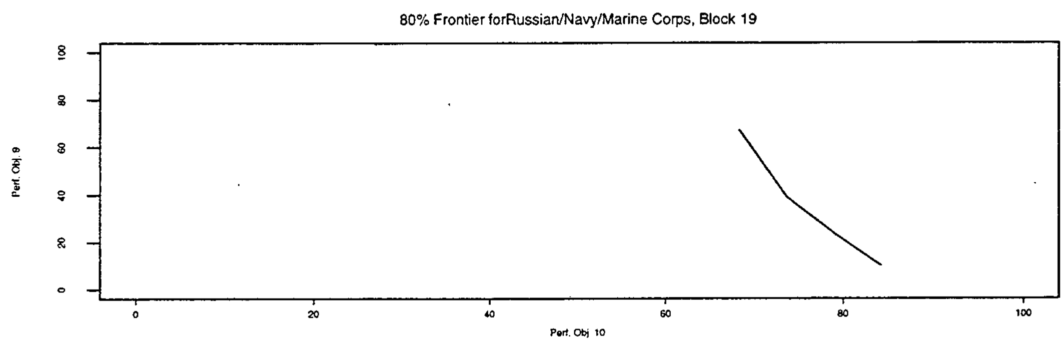


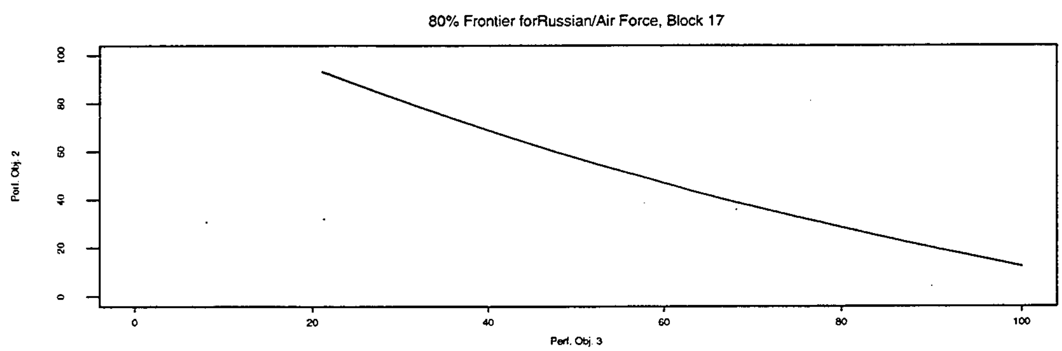
80% Frontier for Vietnamese DLPT_ S

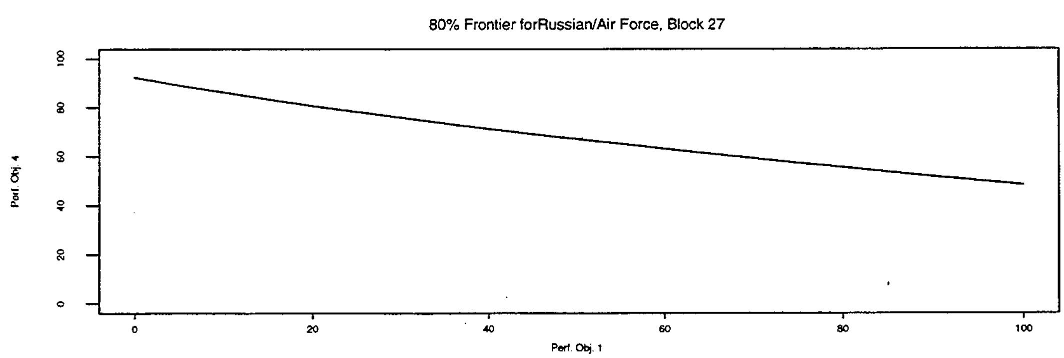


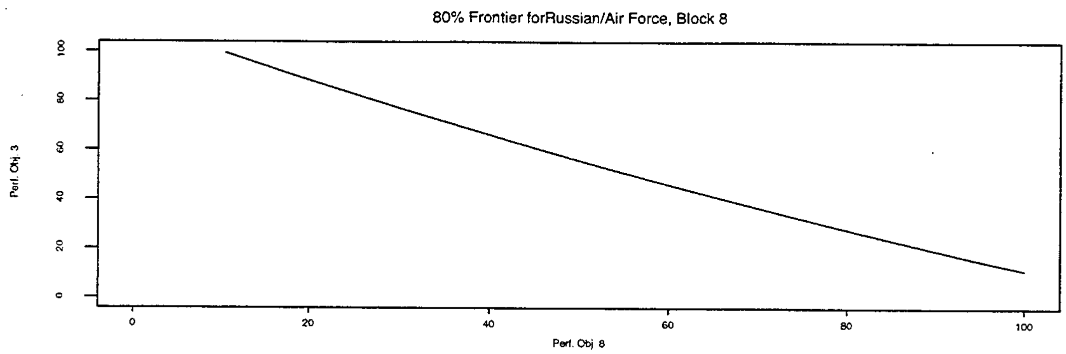
APPENDIX G. FRONTIER GRAPHS FOR GAFB BLOCK TESTS

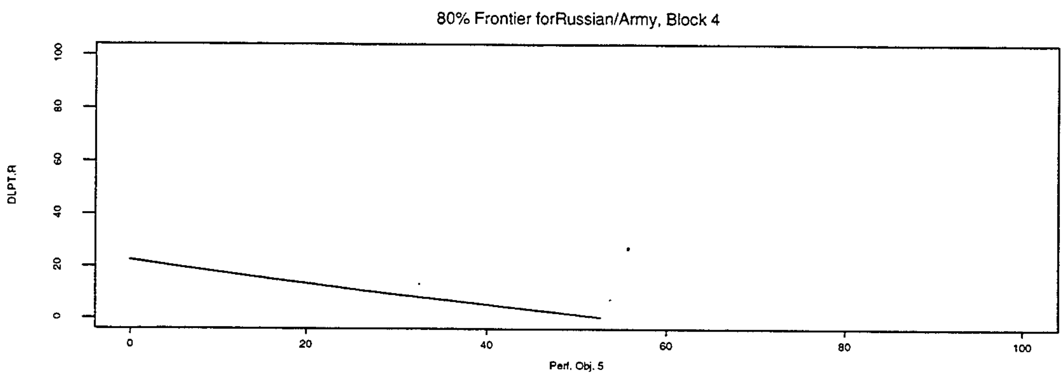









80% Frontier forRussian/Air Force, Block 8
Perf. Obj. 3
Perf. Obj 8
0
20
40
60
80
100


80% Frontier forRussian/Army, Block 4
DLPT.R
Perf. Obj. 5
0
20
40
60
80
100

LIST OF REFERENCES

1. Freund, Rudolf J. and Paul D. Minton, *Regression Methods*, p. 123, Marcel Dekker, Inc., 1979.

2. Hamilton, Lawrence C., *Regression with Graphics*, pp. 30-112, Duxbury Press, 1992.

3. Jackson, G. L. (Ed.) *Language Skill Change Project – Complete Set (LSCP(C)).* Reston, VA: PRC, Inc., 1994. (Available from DLIFLC, ATN: ATFL-ESR, Presidio of Monterey, CA 93944-5006)

4. Jackson, G.L. & Shaw, V.M.W. *Langage Choice and Performance* (Research Report 95-01). Presidio of Monterey, CA: DLIFLC

5. O'Mara, F.E. *Training Approaches For Reducing Student Attrition From Foreign Language Training (LSCP RIII).* Reston,VA: PRC, Inc., 1994

6. O'Mara, F.E, Lett, J.A., Jr., & Alexander, E.E. *The Prediction of Language Learning Success (LSCP RII).* Reston,VA: PRC, Inc., 1994

7. Shaw, V.M.W. & Jackson, G.L. & Lett, J.A., Jr. *The Effects of Length of Service and Prior Language Study at DLI on DLPT Attainment* (Research Report 93-04). Presidio of Monterey, CA: DLIFLC

8. Shaw, V.M.W. & Lett, J.A., Jr. *Relationships of Language Attainment and Age to DLPT Results Among Senior Officer Students in DLIFLC Basic Language Courses* (Research Report 93-03). Presidio of Monterey, CA: DLIFLC

9. S-Plus 4 Guide to Statistics. Seattle, WA: MathSoft, Inc. 1997.

10. SPSS Users' Guide (ver. 7.5) Chicago, IL: SPSS, Inc. 1997.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center...2
   8725 John J. KingmanRd., STE 0944
   Ft. Belvoir, VA 22060-6218

2. Dudley Knox Library...2
   Naval Postgraduate School
   411 Dyer Rd.
   Monterey, CA 93943-5101

3. Samuel Buttrey, Code OR/Sb..2
   Department of Operations Research
   Naval Postgraduate School
   Monterey, CA 93943-5101

4. John Thain..2
   Research and Analysis Division
   Defense Language Institute Foreign Language Center
   Presidio of Monterey, CA 93944-5006

5. John Lett...2
   Research and Analysis Division
   Defense Language Institute Foreign Language Center
   Presidio of Monterey, CA 93944-5006

6. AISO Library..1
   Defense Language Institute Foreign Language Center
   Presidio of Monterey, CA 93944-5006

7. LT Johnna M. Marchant...1
   c/o Judith M. Klingbeil
   33460 Gardenview Dr
   Warren, OR 97053